

(1390 REV. 5-93) US DEPT. OF COMMERCE PATENT &amp; TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER  
105045

**TRANSMITTAL LETTER TO THE  
UNITED STATES  
DESIGNATED/ELECTED OFFICE  
(DO/EO/US) CONCERNING A FILING  
UNDER 35 U.S.C. 371**

U.S. APPLICATION NO.  
(if known, sec 37 C.F.R.1.5)

09/446024

INTERNATIONAL APPLICATION NO.  
PCT/FR98/01442INTERNATIONAL FILING DATE  
July 6, 1998PRIORITY DATE CLAIMED  
July 7, 1997

## TITLE OF INVENTION

Endogenous Retroviral Sequences, Associated with Autoimmune Diseases or with Pregnancy Disorders

## APPLICANT(S) FOR DO/EO/US

Frederic BESEME, Jean-Luc BLOND, Olivier BOUTON, Bernard MANDRAND, Francois MALLET, Herve PERRON

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
  - a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☒ has been transmitted by the International Bureau.
  - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US)
6. ☒ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
  - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☐ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).

**Items 11. to 16. below concern other document(s) or information included:**

11. ☒ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☒ A FIRST preliminary amendment.  
☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A small entity statement.
16. ☒ Other items or information: Sequence Listing.

U.S. APPLICATION NO. (if known, see 37  
C.F.R. 1.5) **09/446024**INTERNATIONAL APPLICATION NO.  
PCT/FR98/01442ATTORNEY'S DOCKET NUMBER  
10504517. ☒ The following fees are submitted:**Basic National fee (37 CFR 1.492(a)(1)-(5)):**

Search Report has been prepared by the EPO or JPO.....\$840.00

International preliminary examination fee paid to USPTO  
(37 CFR 1.482).....\$670.00No international preliminary examination fee paid to USPTO  
(37 CFR 1.482) but international search fee paid to USPTO  
(37 CFR 1.445(a)(2)).....\$690.00Neither international preliminary examination fee (37 CFR  
1.482) nor international search fee (37 CFR 1.445(a)(2))  
paid to USPTO.....\$970.00International preliminary examination fee paid to USPTO  
(37 CFR 1.482) and all claims satisfied provisions of PCT  
Article 33(2)-(4).....\$ 96.00**ENTER APPROPRIATE BASIC FEE AMOUNT =**Surcharge of \$130.00 for furnishing the oath or declaration later than  
☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR  
1.492(e)).

Claims	Number Filed	Number Extra	Rate
Total Claims	20 - 20 =		X \$ 18.00
Independent Claims	3 - 3 =		X \$ 78.00
Multiple dependent claim(s)(if applicable)			+ \$260.00

**TOTAL OF ABOVE CALCULATIONS =**Reduction by 1/2 for filing by small entity, if applicable. Verified Small  
Entity Statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28). -**SUBTOTAL =**Processing fee of \$130.00 for furnishing the English translation later  
than ☐ 20 ☐ 30 month from the earliest claimed priority date (37 CFR  
1.492(f)). +**TOTAL NATIONAL FEE =**Amount to be  
refunded \$

Charged \$

- a. ☒ Check No. 105087 in the amount of \$840.00 to cover the above fees is enclosed.
- b. ☐ Please charge my Deposit Account No. \_\_\_\_\_ in the amount of \$ \_\_\_\_\_ to cover the above fees. A duplicate copy of this sheet is enclosed.
- c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment, to Deposit Account No. 15-0461. A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

OLIFF & BERRIDGE, PLC  
P.O. Box 19928  
Alexandria, Virginia 22320NAME: William P. Berridge  
REGISTRATION NUMBER: 30,024

U.S. APPLICATION NO. (if known, see 37  
C.F.R. 1.5) **09/446024**INTERNATIONAL APPLICATION NO.  
PCT/FR98/01442ATTORNEY'S DOCKET NUMBER  
10504517. ☒ The following fees are submitted:**Basic National fee (37 CFR 1.492(a)(1)-(5)):**

Search Report has been prepared by the EPO or JPO.....\$840.00

International preliminary examination fee paid to USPTO  
(37 CFR 1.482).....\$670.00No international preliminary examination fee paid to USPTO  
(37 CFR 1.482) but international search fee paid to USPTO  
(37 CFR 1.445(a)(2)).....\$690.00Neither international preliminary examination fee (37 CFR  
1.482) nor international search fee (37 CFR 1.445(a)(2))  
paid to USPTO.....\$970.00International preliminary examination fee paid to USPTO  
(37 CFR 1.482) and all claims satisfied provisions of PCT  
Article 33(2)-(4).....\$ 96.00**ENTER APPROPRIATE BASIC FEE AMOUNT =**Surcharge of \$130.00 for furnishing the oath or declaration later than  
☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR  
1.492(e)).

Claims	Number Filed	Number Extra	Rate
Total Claims	20 - 20 =		X \$ 18.00
Independent Claims	3 - 3 =		X \$ 78.00
Multiple dependent claim(s)(if applicable)			+ \$260.00

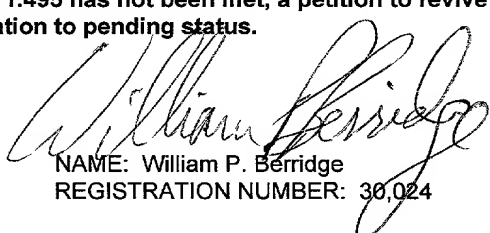
**TOTAL OF ABOVE CALCULATIONS =**Reduction by 1/2 for filing by small entity, if applicable. Verified Small  
Entity Statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28). -**SUBTOTAL =**Processing fee of \$130.00 for furnishing the English translation later  
than ☐ 20 ☐ 30 month from the earliest claimed priority date (37 CFR  
1.492(f)). +**TOTAL NATIONAL FEE =**Amount to be  
refunded \$

Charged \$

- a. ☒ Check No. 105087 in the amount of \$840.00 to cover the above fees is enclosed.
- b. ☐ Please charge my Deposit Account No. \_\_\_\_\_ in the amount of \$\_\_\_\_\_ to cover the above fees. A duplicate copy of this sheet is enclosed.
- c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment, to Deposit Account No. 15-0461. A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

OLIFF & BERRIDGE, PLC  
P.O. Box 19928  
Alexandria, Virginia 22320
  
 NAME: William P. Berridge  
 REGISTRATION NUMBER: 30,024

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of

Frederic BESEME, Jean-Luc BLOND, Olivier  
BOUTON, Bernard MANDRAND, Francois MALLET,  
Herve PERRON

Application No.: New PCT-U.S. National Stage of  
PCT/FR98/01442

Filed: December 16, 1999

Docket No.: 105045

For: ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH  
AUTOIMMUNE DISEASES OR WITH PREGNANCY DISORDERS

PRELIMINARY AMENDMENT

Assistant Commissioner of Patents  
Washington, D. C. 20231

Sir:

Prior to initial examination, please amend the above-identified application as follows:

IN THE TITLE:

Line 1, change "ENDOGENOUS" to --ENDOGENETIC--; and

line 2, change "AND/OR" to --OR--.

IN THE CLAIMS:

Claim 3, line 2, change "either of claims 1 and 2," to --claim 1,--.

Claim 5, lines 1-2, change "either of claims 1 and 4," to --claim 1,--.

Claim 6, lines 1-2, change "either of claims 1 and 4," to --claim 1,--.

Claim 7, lines 5-6, change "any one of claims 1 to 6" to --claim 1,--.

Claim 8, lines 4-6, change "any one of claims 1 to 6, or a nucleic fragment according  
to claim 7." to --claim 1.--.

559221-1203460

Claim 10, lines 5-6, change "any one of claims 1 to 6, or a nucleic fragment according to claim 7." to --claim 1.--.

Claim 15, lines 1-3, change "claims 1 to 6, or of a nucleotide fragment according to claim 7, or of a peptide according to claim 13 or 14," to --claim 1,--.

Claim 16, lines 1-2, change "claims 1 to 6, or of a nucleotide fragment according to claim 7," to --claim 1,--.

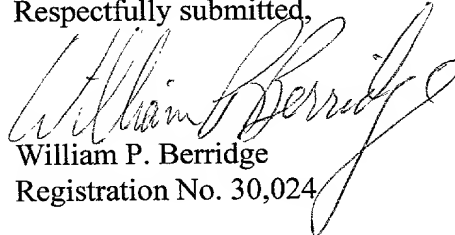
Claim 17, lines 1-2, change "claims 1 to 6, or of a nucleotide fragment according to claim 7," to --claim 1,--.

Claim 20, lines 2-4, change "claims 1 to 6, or a nucleotide fragment according to claim 7, or a peptide according to claim 13 or 14." to --claim 1.--.

#### REMARKS

Claims 1-20 are pending. This Preliminary Amendment corrects typographical errors in the title and eliminates multiple dependent claims. Prompt and favorable examination is respectfully requested.

Respectfully submitted,



William P. Berridge  
Registration No. 30,024

WPB:cas

OLIFF & BERRIDGE, PLC  
P.O. Box 19928  
Alexandria, Virginia 22320  
Telephone: (703) 836-6400

ENDOGENOUS RETROVIRAL SEQUENCES, ASSOCIATED WITH  
AUTOIMMUNE DISEASES AND/OR WITH PREGNANCY DISORDERS

The present invention relates to a new nucleic  
5 material of the endogenous retroviral genomic type,  
various nucleotide fragments comprising it or which are  
obtained from said material, as well as their use as  
marker for at least one autoimmune disease or a  
pathology which is associated with it, a pathological  
10 pregnancy or an unsuccessful pregnancy.

The screening of the cDNA library with the aid  
of the Ppol-MSRV probe (SEQ ID NO: 29) has made it  
possible to detect overlapping clones allowing the  
reconstruction of a putative genomic RNA of  
15 7582 nucleotides. - Reconstructed sequence is under-  
stood to mean the sequence deduced from the alignment  
of the overlapping clones -. This genomic RNA has the  
structure R-U5-gag-pol-env-U3-R. A "blastn" inter-  
rogation on several databases, with the aid of the  
20 reconstructed genome, shows that a large quantity of  
related genomic sequences (DNA) exist in the human  
genome. About 400 sequences have been identified in  
GenBank (cf Figure 3) and more than 200 sequences in  
the EST (Expressed Sequence Tag) library, the majority  
25 as antisense. These sequences are found on several  
chromosomes, in particular chromosomes 5, 7, 14, 16,  
21, 22, X, with a high apparent concentration of LTR on  
the X chromosome.

The reconstructed sequence (mRNA) is integrally  
30 contained inside the genomic clone RG083M05  
(gb AC00064) (9.6 kb), and exhibits 96% similarity with  
two discontinuous regions of this clone which also  
contains repeat regions at each end. The alignment of  
the experimental sequences corresponding to the 5' and  
35 3' regions of the reconstructed genomic RNA with the  
DNA of the RG083M05 clone has made it possible to  
deduce an LTR sequence and to identify elements  
characteristic of retroviruses, in particular those

09/445024  
1999  
09/445024  
1999

involved in reverse transcription, namely the PBS (Primer Binding Site) downstream of the 5' LTR and the PPT (PolyPurine Tract) upstream of the 3' LTR. It is observed that the U3 element is extremely short in comparison with the mammalian type C retroviruses, and comparable in size to the U3 region generally described in the type D retroviruses and the avian retroviruses. The PBS region is homologous to the PBS of the avian retroviruses, suggesting the use of the tRNA<sup>Trp</sup> as primer for the reverse transcription. Consequently, this new family of HERV is called HERV-W (Human Endogenous RetroVirus).

Phylogenetic analysis in the pol region has shown that the HERV-W family is phylogenetically linked to the ERV-9 and RTVL-H families, and therefore belongs to the family of type I endogenous retroviruses. Phylogenetic analysis of the open reading frame (ORF) of env shows that it is closer to the type D simian retroviruses and the avian reticuloendotheliosis retroviruses than type C mammalian retroviruses, suggesting a C/D chimeric genome structure.

The phylogenetic trees, supported by high "bootstrap" values show that the ERV-9 and HERV-W families are derived from two waves of independent insertions. Thus, the active element(s) at the origin of the HERV-W family is (are) different from that (those) from which the ERV-9 family is derived. Furthermore, the PBS of HERV-W probably uses a tRNA<sup>Trp</sup> whereas ERV-9 probably uses a tRNA<sup>Arg</sup>.

Finally, the members of the HERV-W family are expressed in the placenta, whereas the ERV-9 RNAs are not detected in this tissue.

#### BIOLOGICAL FUNCTIONS OF HERV-W

The expression of HERV-W restricted to the placenta and the long reading frame potentially encoding a retroviral envelope make it possible to propose physiological biological functions whose impairment could be associated with pathologies.

0544604.1.21.99

The expression restricted to the placenta suggests that the expression of retroviral and/or nonretroviral genes under the control of the LTRs may be hormone-dependent. These genes may be adjacent, or  
5 under the control of isolated LTRs. A pathology may then result from an aberrant expression following the reactivation of a silent LTR by various factors: viral infection (for example by a member of the Herpesvirus family) or local immune activation. A polymorphism at  
10 the level of the LTRs could also promote these events.

The envelope of HERV-W could play a fusogenic role, in particular at the level of cellular subtypes of the placenta. An immunosuppressive peptide of this envelope could protect the fetus against attack by the  
15 maternal immune system. Finally, by a mechanism of saturation of receptors, the envelope of HERV-W could play a protective role against exogenous retroviral infections. The impairment of local cellular immunity may result from an immunostimulatory signal carried by  
20 the envelope. This effect may be linked to a region carrying a superantigen activity, or to the immunosuppressive region which would become immunostimulatory following either a polymorphism or a dose-effect (overexpression).

25 Verification of these implications and understanding of the consequences linked to an impairment of the biological functions of the endogenous LTRs or the retroviral envelope may lead to the establishment of methods of diagnosis or of monitoring:

- 30 - of states of pathological pregnancy or of unsuccessful pregnancy,  
- of autoimmune diseases such as multiple sclerosis or rheumatoid arthritis.

In accordance with the present invention, there  
35 has been discovered, in the endogenous state, a new nucleic material, stated explicitly and described below, having the organization of a retrovirus, and capable of being correlated with an autoimmune disease,



or a pathology which is associated with it, with a pathological pregnancy or an unsuccessful pregnancy.

The nucleic material according to the present invention, in mRNA form, represents about 8 Kb; it is represented in Figure 1 and is described by SEQ ID NO: 11, and is represented in Figure 2 in the form of genomic DNA.

The expression "of retroviral type" is understood to mean the characteristic according to which the nucleic material considered comprises one or more nucleotide sequences related to the organization of a retrovirus, and/or to its functional or coding sequences.

This reference nucleic material is related to a human endogenous retrovirus, designated by the expression HERV-W. Consequently, it may be obtained by any appropriate technique for screening any library of human DNA, or of placental cDNA, as shown below, in particular with nucleic primers or probes synthesized so as to hybridize with all or part of SEQ ID NO: 11.

The present invention also relates to any nucleic or peptide product, obtained or derived from the reference nucleic material, according to SEQ ID NO: 11.

And finally, the invention relates to the various correlations which may be made between the abovementioned nucleic material, and/or its derived products, with any autoimmune disease and/or a pathology which is associated with it, as well as with cases of pathological pregnancy or of unsuccessful pregnancy.

"Autoimmune" is understood to mean in particular:

- multiple sclerosis
- rheumatoid arthritis
- disseminated lupus erythematosus
- insulin-dependent diabetes
- and/or pathologies which are associated with them.

The present invention relates, first of all, to a nucleic material of the retroviral genomic type, in isolated or purified state, at least partially functional or nonfunctional.

5 This material is characterized in that its genome comprises a reference nucleotide sequence chosen from the group including the sequences SEQ ID NOs: 1 to 15, their complementary sequences, and their equivalent sequences, in particular the nucleotide sequences  
10 exhibiting, for any sequence of 100 contiguous monomers, at least 50% and preferably at least 70%, for example at least 90% homology with respectively said sequences SEQ ID NOs: 1 to 15.

This material is also characterized in that its  
15 genome comprises a reference nucleotide sequence, encoding any polypeptide exhibiting, for any contiguous sequence of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence capable of being encoded by at least a  
20 functional part of the reference nucleotide sequence as defined above.

In particular, this material comprises a nucleic fragment inserted between two sequences corresponding respectively to the LTR region and to the  
25 gag gene for the retroviral genomic structure, in particular a nucleic fragment consisting of or comprising the sequence SEQ ID NO: 12.

The invention also relates to a nucleic material of the subgenomic retroviral type, consisting  
30 of a nucleotide sequence identical to SEQ ID NO: 11, with a deletion as exemplified by the clones cl.PH74 (SEQ ID NO: 7), cl.PH7 (SEQ ID NO: 8) and cl.Pi5T (SEQ ID NO: 9), this deletion resulting or otherwise from a splicing strategy.

35 The above-defined nucleic material comprises at least one functional nucleotide sequence encoding at least one retroviral protein, and/or at least one regulatory nucleotide sequence.

00445024-24699

Next, the invention relates to any nucleotide fragment of at least 100 bases, comprising a nucleotide sequence chosen from the group comprising:

5 a) all the nucleotide sequences, partial and complete, of a nucleic material as defined above

b) all the nucleotide sequences, partial and complete, of a clone chosen from the group including the clones:

- 10 - cl.6A2 (SEQ ID NO: 1)
- cl.6A1 (SEQ ID NO: 2)
- cl.7A16 (SEQ ID NO: 3)
- cl.Pi22 (SEQ ID NO: 4)
- cl.24.4 (SEQ ID NO: 5)
- cl.C4C5 (SEQ ID NO: 6)
- 15 - cl.PH74 (SEQ ID NO: 7)
- cl.PH7 (SEQ ID NO: 8)
- cl.Pi5T (SEQ ID NO: 9)
- cl.44.4 (SEQ ID NO: 10)
- HERV-W (SEQ ID NO: 11)
- 20 - cl.6A5 (SEQ ID NO: 12)
- cl.7A20 (SEQ ID NO: 13)
- cl.7A21 (SEQ ID NO: 14)
- LTR (SEQ ID NO: 15)

25 c) the sequences which are respectively complementary to the sequences according to a) and b)

d) the sequences which are respectively equivalent to the sequences according to a) to c), in particular the nucleotide sequences exhibiting, for any sequence of 100 contiguous monomers, at least 50%, and  
30 preferably at least 70%, or even better at least 80%, for example at least 90% homology with the sequences a) to c).

The invention also relates to any nucleic probe for the detection of a nucleic material, inserted or  
35 otherwise into a nucleic acid, characterized in that it is capable of hybridizing specifically with a nucleic material, as defined above.

Such a probe comprises a marker or otherwise.

The invention also relates to a nucleic primer for the amplification by polymerization of an RNA or of a DNA, characterized in that it comprises a nucleotide sequence capable of hybridizing specifically with a nucleic material or a nucleic fragment, as defined above.

By way of example, a nucleic probe or nucleic primer according to the invention is characterized in that it consists of a nucleotide sequence chosen from the group including SEQ ID NOS: 16 to 28.

The invention also relates to any RNA or DNA, and in particular a replication vector, comprising a nucleotide fragment, as defined above.

The invention also relates to any peptide encoded by any open reading frame belonging to a nucleotide fragment, as defined above, in particular polypeptide, for example oligopeptide forming an antigenic determinant recognized by sera from patients affected by an autoimmune disease, or a pathology which is associated with it, or from patients having a pathological pregnancy or an unsuccessful pregnancy.

By way of example, this polypeptide is encoded by a nucleotide fragment comprising an open reading frame encoding one or more retroviral ENV proteins.

Finally, the invention relates to:

- the use of a nucleic material, or of a nucleotide fragment, or of a peptide defined above, as previously defined, as molecular marker for an autoimmune disease or for a pathology which is associated with it, for pathological pregnancy or unsuccessful pregnancy;

- the use of a nucleic material, or of a nucleotide fragment, as defined above, as chromosomal marker for susceptibility to an autoimmune disease or for a pathology which is associated with it, or for a risk of a pathological pregnancy or of an unsuccessful pregnancy;

- the use of a nucleic material, or of a nucleotide fragment, as defined above, as proximity

09446024 "12169  
65921" 4209460

marker for a gene for susceptibility to an autoimmune disease or to a pathology which is associated with it, or to a risk of a pathological pregnancy or of an unsuccessful pregnancy.

5           The invention also relates to a method for the molecular labeling of an autoimmune disease or of a pathology which is associated with it, of pathological pregnancy or of unsuccessful pregnancy, characterized in that any nucleotide fragment, as defined above,  
10 either in RNA form or in DNA form, is identified and/or quantified in any biological body material, in particular body fluid.

By way of example, according to such a method, cells expressing a nucleotide fragment, as defined  
15 above, are detected in said biological body material.

The invention relates to a diagnostic and/or therapeutic application of a nucleic material, of a nucleotide fragment or of a peptide defined above, and as such, another subject of the invention is a  
20 diagnostic composition or a therapeutic composition comprising said material, said fragment or said peptide.

Before detailing the invention, various terms used in the description and the claims are now defined:

25           - human virus is understood to mean a virus capable of infecting or of being harbored by a human being,

          - taking into account all the natural or induced variations and/or recombinations which may be  
30 encountered in the practical implementation of the present invention, the subjects thereof, defined above and in the claims, have been expressed comprising the equivalents or derivatives of the different biological materials defined below, in particular the homologous  
35 nucleotide or peptide sequences,

          - the variant of a virus or of a pathogenic and/or infective agent according to the invention comprises at least one antigen recognized by at least one antibody directed against at least one corres-

ponding antigen of said virus and/or of said pathogenic and/or infective agent, and/or a genome of which any part is detected by at least one hybridization probe, and/or at least one nucleotide amplification primer  
5 specific for said virus and/or pathogenic and/or infective agent, in particular a genome belonging to the HERV-W family, under determined hybridization conditions well known to persons skilled in the art,

- according to the invention, a nucleotide  
10 fragment or an oligonucleotide or a polynucleotide is a stretch of monomers, or a biopolymer, characterized by the sequence, informational or otherwise, of the natural nucleic acids, capable of hybridizing with any other nucleotide fragment under predetermined conditions, it being possible for the stretch to contain  
15 monomers of different chemical structures and to be obtained from a natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis; a nucleotide fragment may be identical to a genomic  
20 fragment of an element of the HERV-W family considered by the present invention, in particular a gene for the latter, for example pol or env in the case of said element;

- thus, a monomer may be a natural nucleotide  
25 of a nucleic acid, whose constituent elements are a sugar, a phosphate group and a nitrogen base; in RNA, the sugar is ribose, in DNA, the sugar is 2-deoxyribose; depending on whether DNA or RNA is involved, the nitrogen base is chosen from adenine,  
30 guanine, uracil, cytosine, thymine; or the nucleotide may be modified in at least one of the three constituent elements; by way of example, the modification may take place at the level of the bases, generating modified bases such as inosine, 5-methyl-  
35 deoxycytidine, deoxyuridine, 5-(dimethylamino)deoxyuridine, 2,6-diaminopurine, 5-bromodeoxyuridine and any other modified base promoting hybridization; at the level of the sugar, the modification may consist in the replacement of at least one deoxyribose with a

polyamide, and at the level of the phosphate group, the modification may consist in its replacement with esters, in particular chosen from diphosphate, alkyl and arylphosphonate and phosphorothioate esters,

5           - "functional" is understood to mean the characteristic according to which a nucleotide sequence, a nucleic material or a nucleotide fragment comprises an "an informational sequence",

10           - "informational sequence" is understood to mean any ordered sequence of monomers whose chemical nature and the order in a reference direction, constitute or otherwise a functional information of the same quality as that of the natural nucleic acids, for example a reading frame encoding a protein, a  
15           regulatory sequence, a splicing site or a recombination site,

            - hybridization is understood to mean the process during which, under appropriate operating, in particular, stringency, conditions, two nucleotide  
20           fragments, having sufficiently complementary sequences, pair to form a complex, in particular double or triple, structure, preferably in the form of a helix,

            - a probe comprises a nucleotide fragment synthesized in particular by the chemical or polymerization route, or obtained by enzymatic digestion or cleavage of a longer nucleotide fragment, comprising at least six monomers, advantageously from 10 to 100 monomers, preferably 10 to 30 monomers, and possessing a hybridization specificity under determined  
25           conditions; preferably, a probe possessing less than 10 monomers is not used alone, but is used in the presence of other probes equally short in size or otherwise; under certain specific conditions, it may be useful to use probes larger than 100 monomers in size;  
30           a probe may in particular be used for diagnostic purposes and it will include for example capture and/or detection probes,

            - the capture probe may be immobilized on a solid support by any appropriate means, that is to say

directly or indirectly, for example by covalence or by passive adsorption,

- the detection probe may be labeled by means of a marker chosen in particular from radioactive isotopes, enzymes particularly chosen from peroxidase and alkaline phosphatase and those capable of hydrolyzing a chromogenic, fluorogenic or luminescent substrate, chromophoric chemical compounds, chromogenic, fluorogenic or luminescent compounds, nucleotide base analogs, and biotin,

- the probes used for diagnostic purposes of the invention may be used in all the hybridization techniques known to persons skilled in the art, and in particular the techniques termed "DOT-BLOT", "SOUTHERN BLOT", "NORTHERN BLOT" which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as target, the SANDWICH technique; advantageously, the SANDWICH technique is used in the present invention, comprising a specific capture probe and/or a specific detection probe, it being understood that the capture probe and the detection probe must have a nucleotide sequence which is at least partially different,

- any probe according to the present invention may hybridize in vivo or in vitro with RNA and/or with DNA, to block the phenomena of replication, in particular translation and/or transcription, and/or to degrade said DNA and/or RNA,

- a primer is a probe comprising at least six monomers, and advantageously from 10 to 30 monomers, possessing a hybridization specificity under determined conditions, for the initiation of an enzymatic polymerization, for example in an amplification technique such as PCR (Polymerase Chain Reaction), in an extension method such as sequencing, in a reverse transcription method and the like,

- two nucleotide or peptide sequences are said to be equivalent or derived from each other, or relative to a reference sequence, if functionally the corresponding biopolymers may play substantially the

09446024 12199  
669221 42094450



same role, without being identical, in relation to the application or use considered, or in the technique in which they are used; in particular equivalent are two sequences obtained because of the natural variability within the same individual, or the natural diversity from one individual to another within the same species, in particular spontaneous mutation of the species from which they were identified, or induced mutation, as well as two homologous sequences, the homology being defined below,

- "variability" is understood to mean any modification, spontaneous or induced, of a sequence, in particular by substitution, and/or insertion, and/or deletion of nucleotides and/or of nucleotide fragments, and/or extension and/or shortening of the sequence at at least one of the ends; an unnatural variability may result from the genetic engineering techniques used, for example from the choice of the synthetic primers, degenerate or otherwise, selected for amplifying a nucleic acid; this variability may result in modifications of any starting sequence, considered as reference, and which may be expressed by a degree of homology relative to said reference sequence,

- homology characterizes the degree of identity of two nucleotide or peptide fragments compared; it is measured by the percentage identity which is in particular determined by direct comparison of nucleotide or peptide sequences, relative to reference nucleotide or peptide sequences,

- this percentage identity was specifically determined for the nucleotide fragments, in particular clones within the present invention, and obtained from the same individual; by way of nonlimiting example, the lowest percentage identity observed between the different clones from the same individual (cf SEQ ID NOs: 13 and 14) is at least 90% and the lowest percentage identity observed between the different clones of two individuals is at least 80%,

09445024.12054460

- any nucleotide fragment is said to be equivalent to or derived from a reference fragment if it exhibits a nucleotide sequence equivalent to the sequence of the reference fragment; according to the  
5 above definition, particularly equivalent to a reference nucleotide fragment are:

(a) any fragment capable of at least partially hybridizing with the complement of the reference fragment,

10 (b) any fragment whose alignment with the reference fragment leads to identical contiguous bases being identified in a larger number than with any other fragment obtained from another taxonomic group,

(c) any fragment resulting or capable of  
15 resulting from the natural variability within the same individual, and from the natural diversity from one individual to another within the same species, from which it is obtained,

(d) any fragment capable of resulting from  
20 genetic engineering techniques applied to the reference fragment,

(e) any fragment, containing at least eight contiguous nucleotides, encoding a peptide homologous or identical to the peptide encoded by the reference  
25 fragment,

(f) any fragment different from the reference fragment by insertion, deletion, substitution of at least one monomer, extension, or shortening at at least one of its ends; for example, any fragment corresponding to the reference fragment, flanked at at least  
30 one of its ends by a nucleotide sequence not encoding a polypeptide,

- partial or complete nucleotide sequence of a reference nucleic material is also understood to mean  
35 any sequence associated by co-encapsulation, or by coexpression, or recombined with said reference nucleic material,

- polypeptide is understood to mean in particular any peptide of at least two amino acids, in

0944604 2169  
"SECRET" 42094450

particular oligopeptide or a protein, extracted, separated or substantially isolated or synthesized, through the intervention of human hands, in particular those obtained by chemical synthesis, or by expression  
5 in a recombinant organism,

- polypeptide partially encoded by a nucleotide fragment is understood to mean a polypeptide having at least three amino acids encoded by at least nine contiguous monomers contained in said nucleotide  
10 fragment,

- an amino acid is said to be analogous to another amino acid when their respective physico-chemical characteristics, such as polarity, hydrophobicity and/or basicity, and/or acidity, and/or  
15 neutrality, are substantially the same; thus, a leucine is analogous to an isoleucine,

- any polypeptide is said to be equivalent to or derived from a reference polypeptide if the compared polypeptides have substantially the same properties, and in particular the same antigenic, immunological, enzymological and/or molecular recognition properties; particularly equivalent to a reference polypeptide is:  
20

(a) any polypeptide possessing a sequence in which at least one amino acid has been substituted with an analogous amino acid;  
25

(b) any polypeptide having an equivalent peptide sequence obtained by natural or induced variation of said reference polypeptide, and/or of the nucleotide fragment encoding said polypeptide,  
30

(c) a mimotope of said reference polypeptide,

(d) any polypeptide in whose sequence one or more amino acids of the L series are replaced by an amino acid of the D series, and vice versa,

(e) any polypeptide into whose sequence a  
35 modification of the side chains of the amino acids has been introduced, such as for example an acetylation of the amine functions, a carboxylation of the thiol functions, an esterification of the carboxyl functions,

SECRET " 12094460

(f) any polypeptide in whose sequence one or more peptide bonds have been modified, such as for example the carba, retro, inverse, retro-inverse, reduced and methyleneoxy bonds,

5 (g) any polypeptide of which at least one antigen is recognized by an antibody directed against a reference polypeptide,

- the percentage identity characterizing the homology between two compared peptide fragments is, according to the present invention, at least 80% and preferably at least 90%.

The expressions relating to order which are used in the present description and the claims, such as "first nucleotide sequence" are not selected to express a particular order, but to define the invention more clearly.

Detection of a substance or agent is understood to mean hereinafter both an identification and a quantification, or a separation or isolation of said substance or of said agent.

The invention will be understood more clearly upon reading the detailed description which follows, made with reference to the appended figures in which:

- Figure 1 represents, on the one hand, the organization of the endogenous retroviral material discovered according to the present invention, in the form of a putative genomic mRNA, and, on the other hand, the location of the clones used according to the present invention, relative to this organization; the scales for length are expressed in Kb; the flanking regions (5' UTR and 3' UTR) are indicated in hatched boxes; the regions repeated in these two flanking regions are indicated by black arrows; the regions corresponding to the gag, pol and env genes are indicated in black, white and gray respectively; the position of the Ppol-MSRV probe is indicated;

- Figure 2 represents a possibility of genetic organization (DNA), illustrated by the clone RG083M05, and a splicing strategy linking to this sequence, the

0946024 24693  
SECRET

experimental clones (mRNA); this figure also shows the splicing sites observed with reference to the retroviral organization; additionally indicated in this figure are:

5           the location of the probes used (Pgag-LB19, Ppro-E, Ppol-MSRV and Penv-C15);  
          the splice donor sites (DS1 and DS2) and acceptor sites (AS1 to AS3);  
          the sequences obtained from the clone RG083M05,  
10   in the lower-case boxes, and the sequences derived from experimental placental clones (mRNA), in the upper-case boxes;

          the putative ORFs (ORF1, ORF2 and ORF3); and  
          an insert of 2 Kb present in DNA form but not  
15   detected in RNA form, represented in the form of vertical hatches.

The other conventions used in this figure are the same as those for Figure 1.

- Figure 3 gives a representation of genomic  
20   (DNA) clones corresponding to the isolated cDNA clones; indicated in this figure are:

          the percentage similarity with respect to the reconstructed genomic RNA (Recons RNA);

          the presence of repeat sequences at each end of  
25   these genomes (repeats); and

          the presence and the size of the open reading frames (ORFs).

- Figure 4 represents a phylogenetic analysis identifying the HERV-W family.

30           - Figure 5 represents the alignment of the 5' and 3' flanking regions of the clone RG083M05 with the terminal 5' and/or 3' regions of some placental clones; the CAAC tandem flanking the 3' and 5' LTRs is doubly underlined under the DNA sequences, the consensus LTR  
35   sequence of 783 bp (base pairs) is indicated under the alignment; the PPT upstream of the 5' end of LTR and the PBS downstream of the 3' end of LTR are indicated; the U3R and U5 regions are indicated; the sites corresponding to the binding of the transcription

factor are underlined and numbered from 1 to 6; the region -73 to 284 corresponds to the sequence evaluated in "CAT assay"; \* corresponds to putative sites for "capping"; [polyA] indicates the polyadenylation  
5 signal.

- Figure 6 represents a putative sequence of a HERV-W envelope polypeptide (ORF1) obtained from 3 different placental cDNA clones; the leader peptide (L), the surface protein (SU) and the transmembrane  
10 protein (TM) are indicated by arrows; the hydrophobic fusion peptide and the transmembrane carboxy region are underlined by a single line and a double line, respectively; the immunosuppression region is indicated in italics; the potential glycosylation sites are  
15 indicated by dots; the divergent amino acids are indicated on the bottom line; Figure 6 also presents the open reading frames corresponding to ORF2 and ORF3 as described in Figure 2, and more particularly their homologies with the retroviral regulatory genes.

The nucleic material previously presented explicitly was discovered and characterized at the end of the experimental protocol described below, it being understood that this protocol cannot limit the scope of the present invention and of the accompanying claims.

#### 25 **Example 1**

##### **Isolation and sequencing of overlapping cDNA fragments**

The information relating to the organization of HERV-W were obtained by testing a placental cDNA  
30 library (Clontech cat#HL5014a) with the probes Ppol-MSRV (SEQ ID NO: 29) and Penv-C15 (SEQ ID NO: 31) (cf Example 8), and then performing a "gene walking" technique with the aid of the new sequences obtained. The experiments were carried out with reference to the  
35 recommendations of the supplier of the library. PCR amplifications on DNA were also exploited in order to understand this organization.

A number of clones were selected and sequenced, cf Figure 1:

- clone cl.6A2 (SEQ ID NO: 1): untranslated 5' region of HERV-W and part of gag
- clone cl.6A1 (SEQ ID NO: 2): gag and part of pol
- 5 - clone cl.7A16 (SEQ ID NO: 3): 3' region of pol
- clone cl.Pi22 (SEQ ID NO: 4): 3' region of pol and beginning of env
- clone cl.24.4 (SEQ ID NO: 5): spliced RNA
- 10 comprising part of the untranslated 5' region of HERV-W, the end of pol and the 5' region of env
- clone cl.C4C5. (SEQ ID NO: 6): end of env and untranslated 3' region of HERV-W
- clone cl.PH74 (SEQ ID NO: 7): subgenomic RNA:
- 15 untranslated 5' region of HERV-W, end of pol, env and untranslated 3' region of HERV-W
- clone cl.PH7 (SEQ ID NO: 8): multispliced RNA: untranslated 5' region of HERV-W, end of env and untranslated 3' region of HERV-W.
- 20 - clone cl.Pi5T (SEQ ID NO: 9): partial pol gene and U3-R region
- clone cl.44.4 (SEQ ID NO: 10): R-U5 region, gag gene and partial pol gene.

With the aid of these clones, by carrying out

25 sequence alignments, a model of complete sequence of HERV-W was produced. The spliced RNAs were identified as well as the potential splice donor and acceptor sites. This set of information is shown in Figure 2. Through a study of similarity with existing

30 retroviruses, the LTR, gag, pol and env entities were defined.

The putative genetic organization of HERV-W in RNA form is the following (SEQ ID NO: 11):

gene 1..7582

35 location of the clones on the reconstructed genomic RNA sequence

cl.6A2 (1321 bp) 1-1325;

cl.PH74 (535+2229= 2764 bp) 72-606 and 5353-7582;

054400412460450

cl.24.4 (491+1457= 1948 bp); 115-606 and  
 5353-6810;  
 cl.44.4 (2372 bp) 115-2496;  
 cl.PH7 (369+297= 666 bp) 237-606 and 7017-  
 5 7313;  
 cl.6A1 (2938 bp) 586-3559.;  
 cl.Pi5T (2785+566= 3351 bp) 2747-5557 and  
 7017-7582;  
 cl.7A16 (1422 bp) 2908-4337;  
 10 cl.Pi22 (317+1689= 2006 bp) 3957-4273 and  
 4476-6168;  
 cl.C4C5 (1116 bp) 6467-7582  
 5' LTR 1..120  
 /note="R of 5' LTR (5' end uncertain"  
 15 121..575  
 /note="U5 of 5' LTR"  
 various 579..596  
 /note="PBS primer binding site for tRNA-W"  
 various 606  
 20 /note="splice junction (splice donor site  
 ATCCAAAGTG-GTGAGTAATA and splice acceptor  
 site CTTTTTTCAG-ATGGGAAACG clone RG083M05,  
 GenBank accession AC000064)"  
 various 5353  
 25 /note="splice acceptor site for ORF1 (env)"  
 various 5560  
 /note="splice donor site"  
 ORF 5581..7194  
 /note="ORF1 env 538 AA"  
 30 /product="envelope"  
 various 7017  
 /note="splice acceptor site for ORF2 and  
 ORF3"  
 ORF 7039..7194  
 35 /note="ORF2 52 AA"  
 ORF 7112..7255  
 /note="ORF3 48 AA"  
 various 7244..7254  
 /note="PPT polypurine tract"

0944624 2209460  
 66922 1209460



3'LTR 7256..7582  
/note--="U3-R of 3' LTR (U3-R junction  
indeterminate)  
various 7563..7569  
5 polyadenylation signal

**Example 2:****Identification of genomic (DNA) clones corresponding to the isolated DNA clones**

A "blastn" interrogation of several databases,  
10 with the aid of the reconstructed genome, shows that a  
large quantity of related sequences exist in the human  
genome. About 400 sequences were identified in GenBank  
and more than 200 sequences in the EST library, and the  
majority as antisense. The 4 sequences most significant  
15 in size and in similarity, illustrated in Figure 3, are  
the following genomic (DNA) clones:

the human clone RG083M05 (gb AC000064) whose  
chromosomal location is 7q21-7q22,

the human clone BAC378 (gb U85196, gb AE000660)  
20 corresponding to the alpha delta locus of the T cell  
receptor, located in 14q11-12,

the human cosmid Q11M15 (gb AF045450) corres-  
ponding to the 21q22.3 region of chromosome 21,

the cosmid U134E6 (embl Z83850) on chromosome  
25 Xq22.

The location of the aligned regions for each of  
the clones is indicated and the affiliation to a  
chromosome is indicated in square brackets. The  
percentage similarity (without broad deletions) between  
30 the 4 sequences and the reconstructed genomic RNA is  
indicated, as well as the presence of repeat sequences  
at each end of the genome and the size of the largest  
reading frames (ORF). Repeat sequences are found at the  
ends of 3 of these clones. The reconstructed sequence  
35 is integrally contained inside the clone RG083M05  
(9.6 Kb) and exhibits a 96% similarity. However, the  
clone RG083M05 exhibits an insert of 2 Kb situated  
immediately downstream of the untranslated 5' region  
(5' UTR). This insert is also found in two other

genomic clones which exhibit a deletion of 2.3 Kb immediately upstream of the untranslated 3' region (3' UTR). No clone contains the three functional reading frames (ORFs) gag, pol and env. The clone  
5 RG083M05 shows an ORF of 538 amino acids (AA) corresponding to a whole envelope. The cosmid Q11M15 contains two large contiguous ORFs of 413 AA (frame 0) and 305 AA (frame +1) corresponding to a truncated pol polyprotein.

10 **Example 3**

**Phylogenetic analysis**

A phylogenetic analysis was carried out at the level of the nucleic acids on 11 different subregions of the reconstructed genomic RNA, and at the protein  
15 level on 2 different subregions of env. All the trees obtained exhibit the same topology regardless of the region studied. This is illustrated in Figure 4 at the level of the nucleic acids in the most conserved LTR and pol regions between the sequences obtained and  
20 ERV-9 and RTLHV-H. The trees clearly show that the experimental sequences describe a new family distinct from ERV-9 and very distinct from RTLHV-H as underlined by the "bootstrap" analysis. These sequences are found on several chromosomes, in particular chromosomes 5, 7,  
25 14, 16, 21, 22 and X with a high apparent concentration of LTR on the X chromosome.

Comparison at the protein level between the most conserved regions of the retroviral env proteins shows that the HERV-W family is closer to the type D  
30 simian retroviruses and the avian reticuloendotheliosis retroviruses than the type C mammalian retroviruses.

This suggests a C/D chimeric genomic structure.

**Example 4**

**Identification of the LTR, PPT and PBS elements**

35 The reconstructed sequence (RNA) is integrally contained inside the genomic clone RG083M05 (9.6 Kb) and exhibits a 96% similarity with two discontinuous regions of this clone which also contains repeat regions at each end. The alignment of the experimental

sequences corresponding to the 5' and 3' regions of the genomic RNA reconstructed with the DNA of the clone RG083M05 [5' (5-RG-28000-28872) and 3' (3-RG-37500-38314)] made it possible to deduce an LTR sequence and to identify elements characteristic of the retroviruses, in particular those involved in the reverse transcription, namely PBS downstream of the 5' LTR and the PPT upstream of the 3' LTR (cf Figure 5). It is observed that the U3 element is extremely short in comparison with that observed in the mammalian type C retroviruses, and is comparable in size to the U3 region generally described in the type D retroviruses and the avian retroviruses. The region corresponding to bases 2364 to 2720 of the clone cl.PH74 (SEQ ID NO: 7) was amplified by PCR and subcloned into the vector pCAT3 (Promega) in order to carry out the evaluation of the promoter activity. A significant activity was found in HeLa cells by the so-called "CAT assay" method showing the functionality of the promoter sequence of the LTR.

The PBS region is homologous to the PBS of the avian retroviruses.

#### Example 5

#### Genetic organization and regulation of expression

Organization in DNA form

PCR amplifications were carried out on whole HERV-W clones recovered on human genomic library (see Example 1 for the mode of production), using the following oligonucleotide pairs:

U5 4992 (SEQ ID NO: 16), GAG 4619 (SEQ ID NO: 17)  
GAG 4782 (SEQ ID NO: 18), POL 3167 (SEQ ID NO: 19)  
POL 3390 (SEQ ID NO: 20), POL 5144 (SEQ ID NO: 21)  
POL 5145 (SEQ ID NO: 22), U5 4991 (SEQ ID NO: 23).

The PCRs were carried out under the following conditions:

oligonucleotides at the concentration of 0.33 microMolar

TAQ polymerase buffer Boehringer 1X  
0.5 unit of TAQ polymerase Boehringer  
mixture of dNTP at 0.25 mM each  
0.5 mg of human DNA  
5 final volume 100 ml  
PCR conditions (95°C, 5 min) × 1, (95°C, 30 sec  
+ 54°C, 30 sec + 72°C 3 min) × 35.

The PCR products were then deposited on 1% agarose gel to be analyzed after migration. The set of  
10 PCRs gives amplification fragments of the expected size, except for the LTR-4991--gag-4619 PCR which gives a fragment of size greater by about 2 Kb relative to the expected size (deduced from cDNAs from the placental library). The reconstruction of HERV-W in  
15 endogenous DNA form therefore represents an entity of about 10 Kb.

After cloning, sequencing and analysis of the PCR-4992 gag-4619, the presence of a region of insertion is observed between LTR and gag of  
20 SEQ ID NO: 12 (clone cl.6A5). This region does not correspond to an untranslated traditional region of a retrovirus: no  $\psi$  or PBS region.

The products of PCR pol-3390, pol-5144 were also cloned and two of the clones obtained were  
25 sequenced. The result of these sequences is given by the clones cl.7A20 (SEQ ID NO: 13) and cl.7A21 (SEQ ID NO: 14). Comparison of these two nucleotide sequences gives a score of 90% homology for the relevant region, thus showing the variability of HERV-W  
30 in the same individual.

HERV-W in DNA form is proposed in Figure 2.

General organization: transcription process

The various cDNA clones having been obtained, results acquired in PCR on DNA, there is deduced:

35 - a DNA organization of 10 Kb possessing an insertion sequence of 2 Kb between LTR and gag.

The result of PCR on DNA showing the presence of an insert of 2 Kb between the LTR and gag regions suggests that the cDNAs isolated from the placenta are

obtained from the expression of a genome of the RG083M05 type.

- an RNA organization of 8 Kb resulting from a transcription of 10 Kb followed by a splicing between LTR and gag making it possible to restore a continuity FR (Flanking Region) 5' gag, and thus giving an RNA of 8 Kb as identified in Northern blotting.

The probes gag (Pgag-LB19, SEQ ID NO: 30) and protease (Ppro-E, SEQ ID NO: 32) reveal an RNA having a size close to 8 Kb, the probe Penv-C15 (SEQ ID NO: 31) reveals, in addition, an RNA close to 3.1 Kb. Two probes defined in the untranslated 5' region, obtained by screening of the cDNA library reported above (probe P5'-gag-cl.6A2 derived from the clone cl.6A2 and probe P5'-env-cl.24.4 derived from the clone cl.24.4) reveal the preceding two RNAs and an RNA of about 1.3 Kb. This distribution of the RNAs is typical of complex retrovirus transcripts: a genomic RNA encoding gag-pro-pol, a subgenomic RNA encoding the envelope, and one or more multisplliced RNAs potentially encoding regulatory genes.

The half-life of such an RNA (LTR-R-U5-Insertion-GAG-POL-ENV-U3-R-HERV-W) is probably very short, because no RNA of 10 Kb is detected in Northern blotting. By analyzing and comparing sequences, the potential splice donor sites (DS1 and DS2) and acceptor sites were defined and described in Figure 2.

#### **Example 6**

##### **Transcription in healthy tissues**

Various healthy human tissues were tested by the Northern-blot technique (Human Multiple Tissue Northern Blot, Clontech cat# 7760-1), with the aid of the probes Ppol-MSRV (SEQ ID NO: 29), Pgag-LB19 (SEQ ID NO: 30), Penv-C15 (SEQ ID NO: 31), Ppro-E (SEQ ID NO: 32), P5'-gag-cl.6A2 and P5'-env-cl.24.4, labeled as described in Example 1. The experiments were carried out following the recommendations of the manufacturers, and the autoradiographs were exposed for 5 days. Analysis of the results reveals transcription

products only in the placenta, and in none of the other human tissues tested (heart, brain, lungs, liver, skeletal muscle, kidney and pancreas).

Using an RNA Dot-Blot technique (Clontech: Human RNA Master Blot Cat# 7770-1), and using the experimental protocol recommended by the manufacturer, about forty other tissues, including fetal tissues, were tested: only the placenta gives a specific response after hybridization with the probes Pgag-LB19 (SEQ ID NO: 30) and Penv-C15 (SEQ ID NO: 31).

It is observed that a signal is observed in the kidney in RNA Dot-Blot, which is infirmed by the Northern-blot analysis.

#### Example 7

#### Identification of an mRNA encoding an envelope and the means for detecting it specifically

The screening of a placental cDNA library with the aid of a probe defined in the untranslated 5' region made it possible to isolate a cDNA defined by an untranslated 5' region (5' NTR), a splicing junction, a coding sequence, an untranslated 3' region (3' NTR) and a polyadenylated tail, cl.PH74 (SEQ ID NO: 7). This clone corresponds to a spliced RNA encoding an envelope. By comparing sequences between this cDNA and the endogenous HERV-W model proposed according to Figure 2, a splicing junction is identified on the mRNA, a splicing junction placing in continuity the 5' NTR region and the env gene, leading to the production of a spliced subgenomic RNA encoding the envelope gene. This information made it possible to define an oligonucleotide specific for this mRNA by choosing a location situated on the splicing site (Oligo 5307, according to SEQ ID NO: 24).

The identification of this joining region makes it possible to establish a method of discriminating between endogenous retroviral RNA and DNA, using, in a PCR, an oligonucleotide defined on this joining region, in particular an oligonucleotide chosen from the env gene (Oligo 4986, according to SEQ ID NO: 25).

The PCRs were carried out under the following conditions:

oligonucleotides at the concentration of 0.33 microMolar

- 5           TAQ polymerase buffer Boehringer 1X  
          0.5 unit of TAQ polymerase Boehringer  
          mixture of dNTP at 0.25 mM each  
          0.5 mg of human DNA  
          final volume 100 ml

- 10           On 10 different DNAs tested, this type of PCR  
did not make it possible to obtain amplification  
products. On the other hand, on cDNA derived from  
placental RNA or from cells expressing HERV-W, this PCR  
gives an amplification product. This result therefore  
15           confirms the specifically RNA nature of this subgenomic  
fragment.

#### **Example 8**

#### **Identification of coding sequences contained in a specific mRNA**

- 20           The splicing strategy described in Example 5 is  
compatible with the presence of three reading frames  
ORF1 (SEQ ID NO: 33), ORF2 (SEQ ID NO: 34) and ORF3  
(SEQ ID NO: 35) (cf Figure 6).

- The screening of a placental cDNA library made  
25           it possible to isolate a cDNA (SEQ ID NO: 7, cl.PH74)  
defined by an untranslated 5' region (5' NTR), a  
splicing junction, a coding sequence, an untranslated  
3' region (3' NTR) and a polyadenylated tail. The  
coding sequence is 538 amino acids (SEQ ID NO: 33). The  
30           analyses carried out on databanks make it possible to  
identify characteristics of a complete retroviral  
envelope: initiation of translation of an envelope  
polyprotein, of a highly hydrophobic leader peptide of  
about 21 amino acids, of a surface protein SU, of a  
35           transmembrane protein TM. These two protein entities  
exhibit different potential glycosylation sites. An  
immunosuppressive region is identified within the TM  
protein.

09446024.2169

22 bp and 95 bp upstream of the splice acceptor site, two initiation codons were respectively found which were capable of directing the synthesis of 52 AA (ORF2, SEQ ID NO: 34) and of 48 AA (ORF3, SEQ ID NO: 35). ORF2 consists of part of the carboxy-terminal end of env and ORF3 corresponds to a different but overlapping translation.

No significant homology was found by "blast" interrogation. However, an LFASTA interrogation in a sub-databank limited to the Retroviridae, ORF2 and ORF3 showed a percentage identity of 35% with, respectively, Rex of the human and primate lymphotropic T virus, and with Tat of the simian immunodeficiency virus.

#### Example 9

##### Complexity of the HERV-W family

The number of copies present in the human genome of each of the sequences is evaluated by a Dot-Blot technique, with the aid of the probes Pgag-LB19 (SEQ ID NO: 30), Ppro-E (SEQ ID NO: 32) and Penv-C15 (SEQ ID NO: 31).

Each of the probes is denatured and deposited on a Hybond N+ membrane in an amount of 2.5, 5, 10, 25, 50, 100 pg per deposit. 0.5 mg of human DNA are also deposited on the same membrane. The membranes are dried for 2 hours under vacuum at 80°C. The membranes are then hybridized with the deposited probe. The techniques for labeling the probes, for hybridization and for washing the membranes are the same as for the Southern blotting. After autoradiography of the membranes, levels of signal intensity which are proportional to the deposits on the membrane are observed. After cutting out the hybridization zones, scintillation counting is carried out. By comparison between the dilution series for the probe deposited on the membrane and the result obtained with the human DNA, it is possible to evaluate the number of copies per haploid genome of each of the regions covered by the probes:



- the number of endogenous gag is evaluated from 56 to 112 copies (76)

- the number of endogenous protease is evaluated from 166 to 334 copies (260)

5 - the number of endogenous env is evaluated at less than 52 copies (13).

The screening of  $10^6$  clones of a human placental DNA library (Clontech cat# H15014b) made it possible to count 144 clones recognized by the probe Pgag-LB19, and 64 clones recognized by the probe Penv-C15. 13 clones hybridized conjointly with the probes Penv-C15 and Pgag-LB19 were isolated, confirming the presence of several copies of a genome possessing both gag and env, without consideration of  
15 functionality.

The nucleic material, the nucleotide sequences and the peptides or proteins which may be expressed by said materials and sequences may be used to detect, predict, treat and monitor any autoimmune disease, and  
20 the pathologies which are associated with it, as well as in cases of pathological pregnancy or of unsuccessful pregnancy.

Indeed, the objective and experimental data make it possible to link retrovirus and autoimmune  
25 diseases and retrovirus and pregnancy disorders:

(1) common mechanisms are used in the retroviral pathologies and in autoimmune diseases (presence of autoantibodies, of immune complexes, cellular infiltration of certain tissues, neurological  
30 disorders).

(2) pathological disorders comparable to certain autoimmune diseases appear during infections with HIV and HTLV retroviruses (Sjögren syndrome, disseminated lupus erythematosus, rheumatoid arthritis  
35 and the like).

(3) a reverse transcriptase activity was detected and retroviral-type particles were observed in the cell culture supernatants of patients suffering from multiple sclerosis (Perron et al., Res. Virol.

6694604 1369

1989; 140: 551-561/Lancet 1991; 337: 862-863/Res. Virol. 1992; 143: 337-350) or from rheumatoid arthritis.

(4) autoimmune or chronic inflammatory animal pathologies are linked to endogenous retroviruses; some of them are used as animal models of human diseases (insulin-dependent diabetes, disseminated lupus erythematosus).

(5) significant levels of endogenous anti-retrovirus antibodies have been described in the context of autoimmune, systemic or inflammatory diseases; other data of this nature were communicated by several authors at the IVth European meeting on endogenous retroviruses (Uppsala, October 1996). According to Venables (communiqués of the IVth European meeting on endogenous retroviruses, Uppsala, October 1996), a significantly high level of anti-HERV-H antibodies are found during pregnancy but also in the context of various autoimmune disorders such as Sjögren syndrome, disseminated lupus erythematosus or rheumatoid arthritis, without, however, any proof of its direct involvement being provided up until now.

The involvement of the retroviruses in the autoimmune phenomenon remains compatible with the multifactorial character of the autoimmune, systemic or inflammatory diseases which confront genetic, hormonal, environmental and infectious factors.

The particles observed in the cell culture supernatants from patients suffering from multiple sclerosis (Perron et al., Res. Virol. 1989; 140: 551-561/Lancet 1991; 337: 862-863/Res. Virol. 1992; 143: 337-350) or from rheumatoid arthritis (unpublished data) may result from the expression: (i) of an endogenous retrovirus competent for replication, (ii) of several defective endogenous retroviruses cooperating by a phenomenon of transcomplementation or (iii) of an exogenous retrovirus.

All these observations make it possible to use and consider the above-described biological material as

0944524 269



## CLAIMS

1. Nucleic material of the retroviral genomic type, in isolated or purified state, at least partially functional or nonfunctional, whose genome comprises a reference nucleotide sequence chosen from the group including the sequences SEQ ID NOs: 1 to 15, their complementary sequences, and their equivalent sequences, in particular the nucleotide sequences exhibiting, for any sequence of 100 contiguous monomers, at least 70% and preferably at least 90% homology with respectively said sequences SEQ ID NOs: 1 to 15.
2. Nucleic material of the retroviral genomic type, in isolated or purified state, at least partially functional or nonfunctional, whose genome comprises a reference nucleotide sequence, encoding any polypeptide exhibiting, for any contiguous sequence of at least 30 amino acids, at least 80%, and preferably at least 90% homology with a peptide sequence capable of being encoded by at least a functional part of the reference nucleotide sequence according to claim 1.
3. Nucleic material of the retroviral genomic type according to either of claims 1 and 2, comprising a nucleic fragment inserted between two sequences corresponding respectively to the LTR region and to the gag gene for the retroviral genomic structure, in particular a nucleic fragment consisting of or comprising the sequence SEQ ID NO: 12.
4. Nucleic material of the subgenomic retroviral type, consisting of a nucleotide sequence identical to SEQ ID NO: 11, with at least one deletion, such as a sequence chosen from SEQ ID NOs: 7 to 9.
5. Nucleic material according to either of claims 1 and 4, comprising at least one functional nucleotide sequence encoding at least one retroviral protein.

6. Nucleic material according to either of claims 1 and 4, comprising at least one regulatory nucleotide sequence.

7. Nucleotide fragment of at least 100 bases,  
5 comprising a nucleotide sequence chosen from the group comprising:

a) all the nucleotide sequences, partial and complete, of a nucleic material according to any one of claims 1 to 6

10 b) all the nucleotide sequences, partial and complete, of a clone chosen from the group including the clones:

- cl.6A2 (SEQ ID NO: 1)
- cl.6A1 (SEQ ID NO: 2)
- 15 - cl.7A16 (SEQ ID NO: 3)
- cl.Pi22 (SEQ ID NO: 4)
- cl.24.4 (SEQ ID NO: 5)
- cl.C4C5 (SEQ ID NO: 6)
- cl.PH74 (SEQ ID NO: 7)
- 20 - cl.PH7 (SEQ ID NO: 8)
- cl.Pi5T (SEQ ID NO: 9)
- cl.44.4 (SEQ ID NO: 10)
- HERV-W (SEQ ID NO: 11)
- cl.6A5 (SEQ ID NO: 12)
- 25 - cl.7A20 (SEQ ID NO: 13)
- cl.7A21 (SEQ ID NO: 14)
- LTR (SEQ ID NO: 15)

c) the sequences which are respectively complementary to the sequences according to a) and b)

30 d) the sequences which are respectively equivalent to the sequences according to a) to c), in particular the nucleotide sequences exhibiting, for any sequence of 100 contiguous monomers, at least 50%, and preferably at least 70%, for example at least 90%  
35 homology with the sequences a) to c).

8. Nucleic probe for the detection of a nucleic material, inserted or otherwise into a nucleic acid, characterized in that it is capable of hybridizing

specifically with a nucleic material, according to any one of claims 1 to 6, or a nucleic fragment according to claim 7.

9. Probe according to claim 8, characterized in that it comprises a marker.

10. Nucleic primer for the amplification by polymerization of an RNA or of a DNA, characterized in that it comprises a nucleotide sequence capable of hybridizing specifically with a nucleic material according to any one of claims 1 to 6, or a nucleic fragment according to claim 7.

11. Nucleic probe or nucleic primer, characterized in that it consists of a nucleotide sequence chosen from the group including SEQ ID NOs: 16 to 28.

12. RNA or DNA, and in particular replication vector, comprising a nucleotide fragment according to claim 7.

13. Peptide encoded by any open reading frame belonging to a nucleotide fragment, according to claim 7, in particular polypeptide, for example oligopeptide forming an antigenic determinant recognized by sera from patients affected by an autoimmune disease, or a pathology which is associated with it, or from patients having a pathological pregnancy or an unsuccessful pregnancy.

14. Peptide according to claim 13, characterized in that it is encoded by a nucleotide fragment comprising an open reading frame encoding one or more retroviral ENV proteins.

15. Use of a nucleic material according to claims 1 to 6, or of a nucleotide fragment according to claim 7, or of a peptide according to claim 13 or 14, as molecular marker for an autoimmune disease or for a pathology which is associated with it, or for a pathological pregnancy or for an unsuccessful pregnancy.

16. Use of a nucleic material according to claims 1 to 6, or of a nucleotide fragment according to claim 7,

as chromosomal marker for susceptibility to an autoimmune disease or for a pathology which is associated with it, or for a risk of a pathological pregnancy or of an unsuccessful pregnancy.

- 5 17. Use of a nucleic material according to claims 1 to 6, or of a nucleotide fragment according to claim 7, as proximity marker for a gene for susceptibility to an autoimmune disease or to a pathology which is associated with it, or to a risk of a pathological  
10 pregnancy or of an unsuccessful pregnancy.

18. Method for the molecular labeling of an autoimmune disease or of a pathology which is associated with it, of a pathological pregnancy or of an unsuccessful pregnancy, characterized in that any  
15 nucleotide fragment according to claim 7, either in RNA form or in DNA form, is identified and/or quantified in any biological body material, in particular body fluid.

19. Method according to claim 18, characterized in that cells expressing the nucleotide fragment according  
20 to the claim are detected in said biological body material.

20. Diagnostic or therapeutic composition comprising a nucleic material according to claims 1 to 6, or a nucleotide fragment according to claim 7, or a  
25 peptide according to claim 13 or 14.

FIG 1

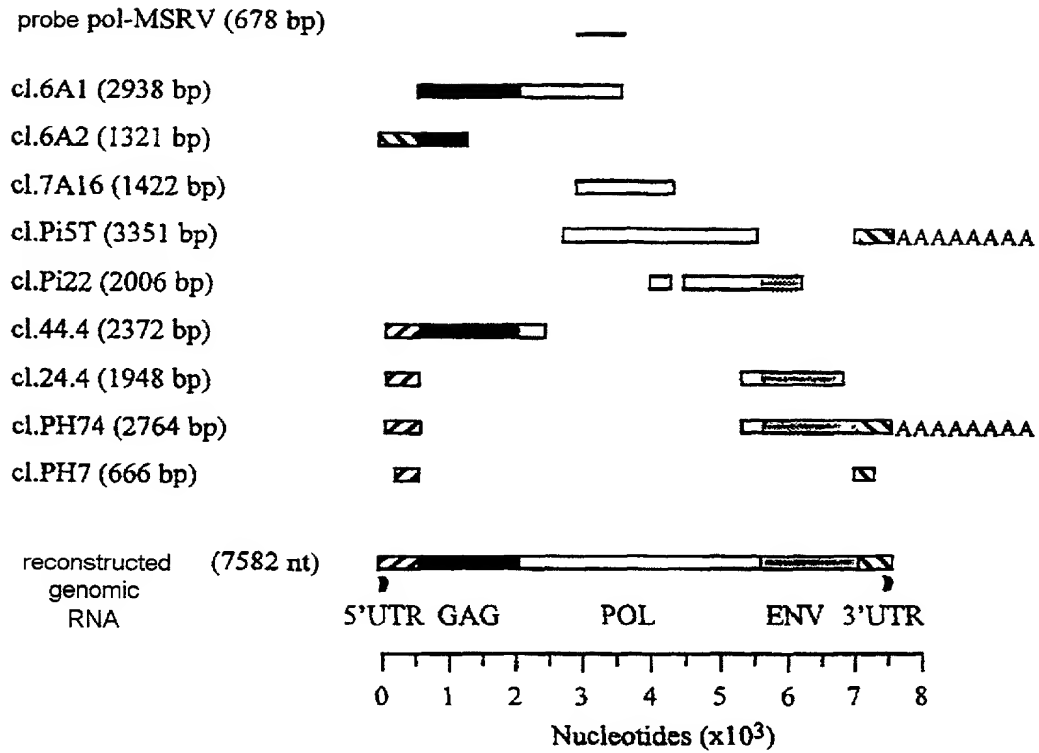




FIG 2

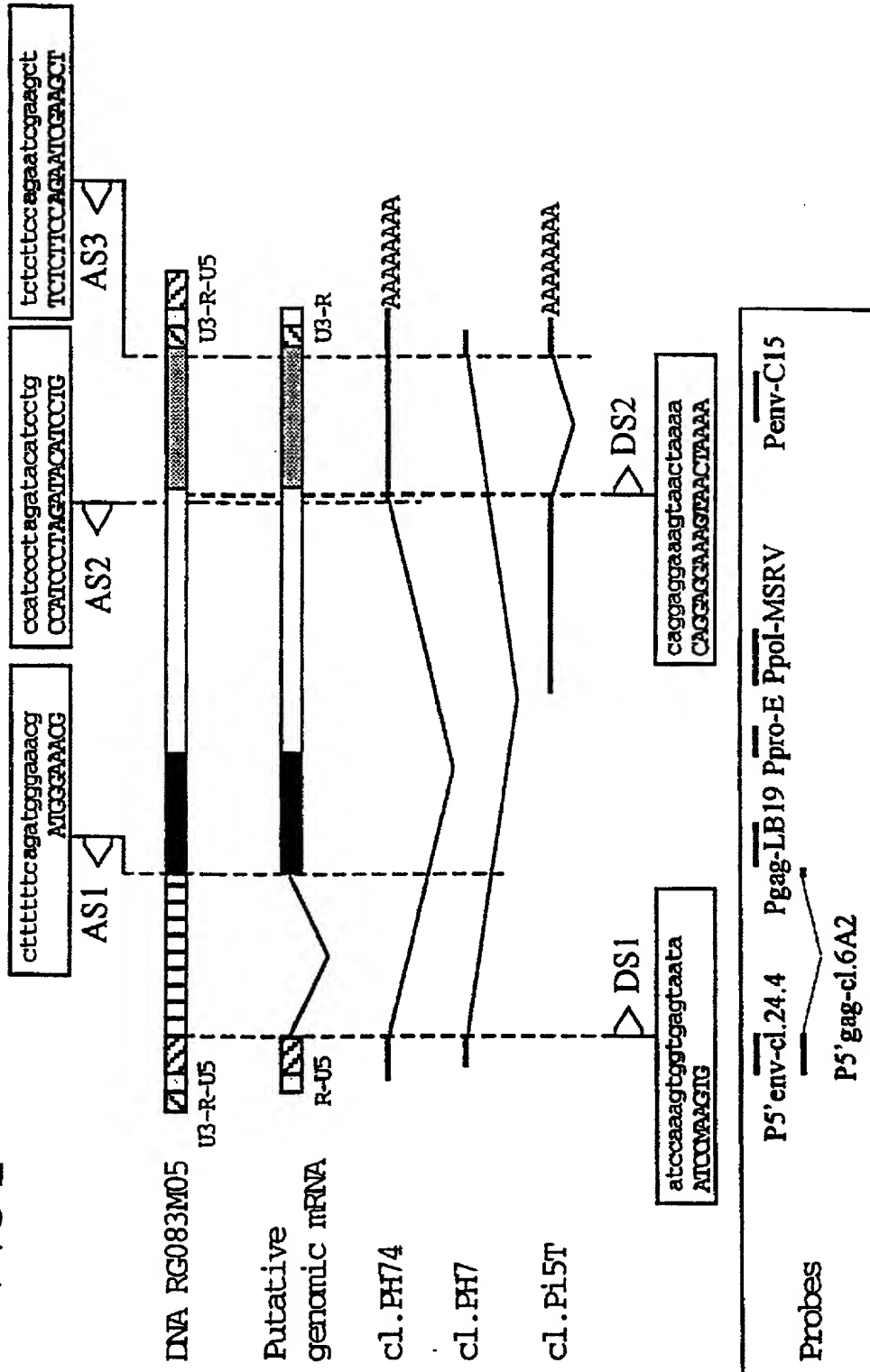


FIG 3

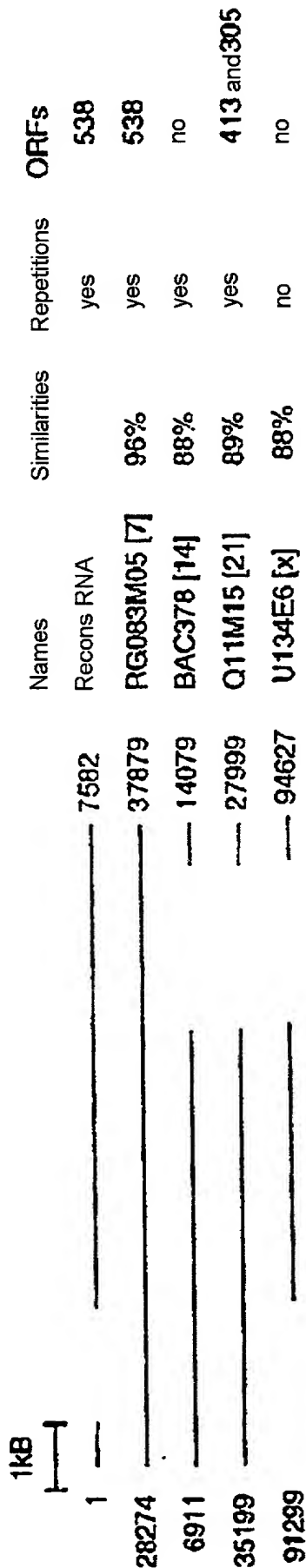


FIG 4A

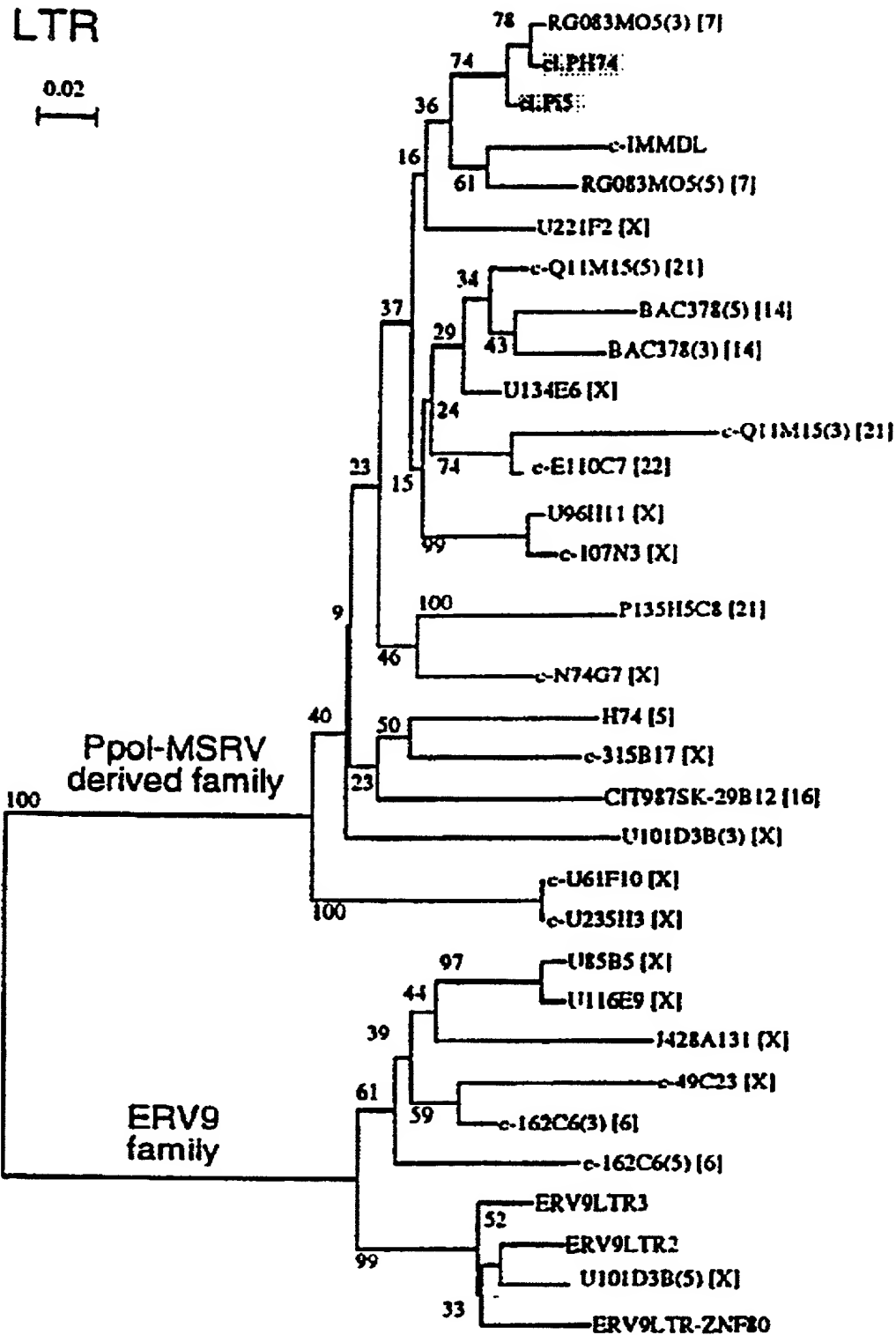


FIG 4 B

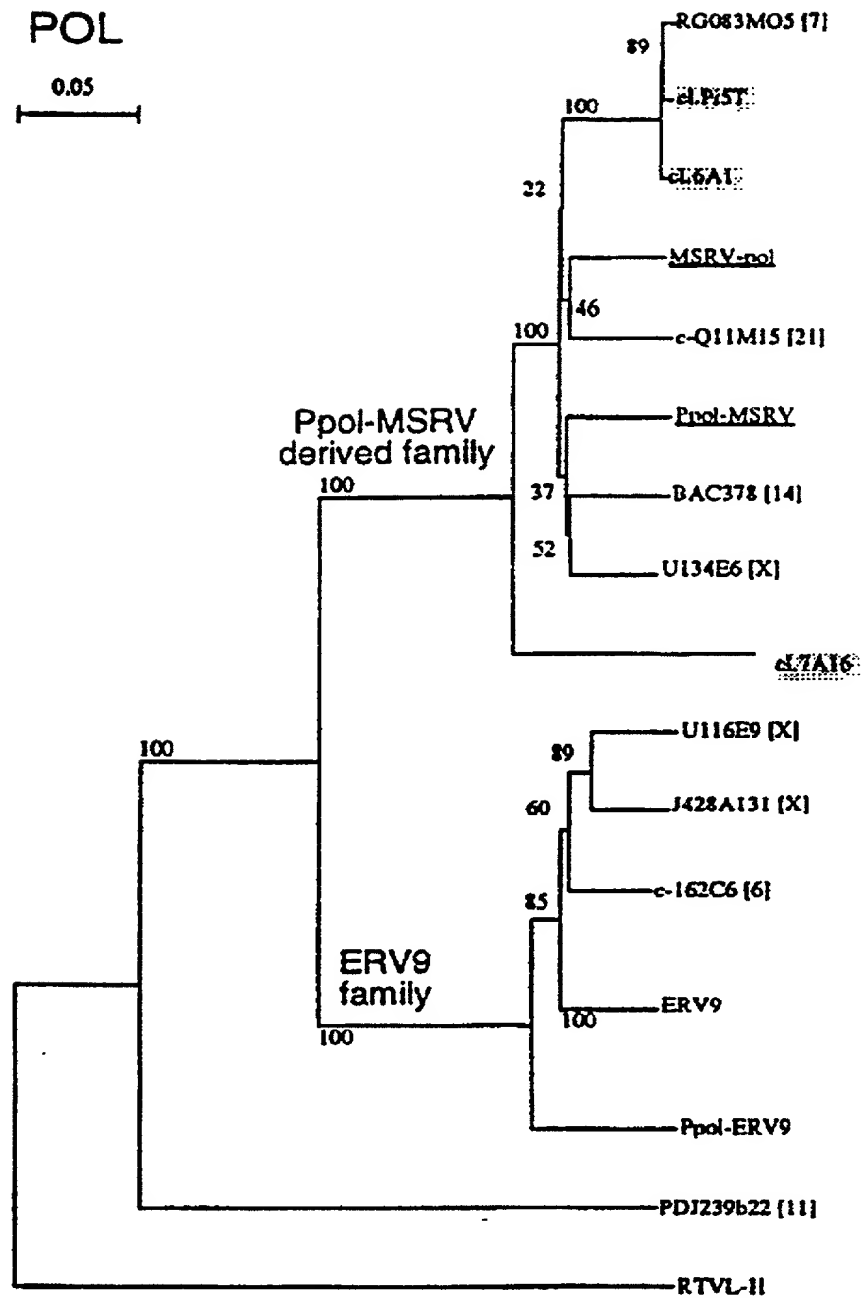
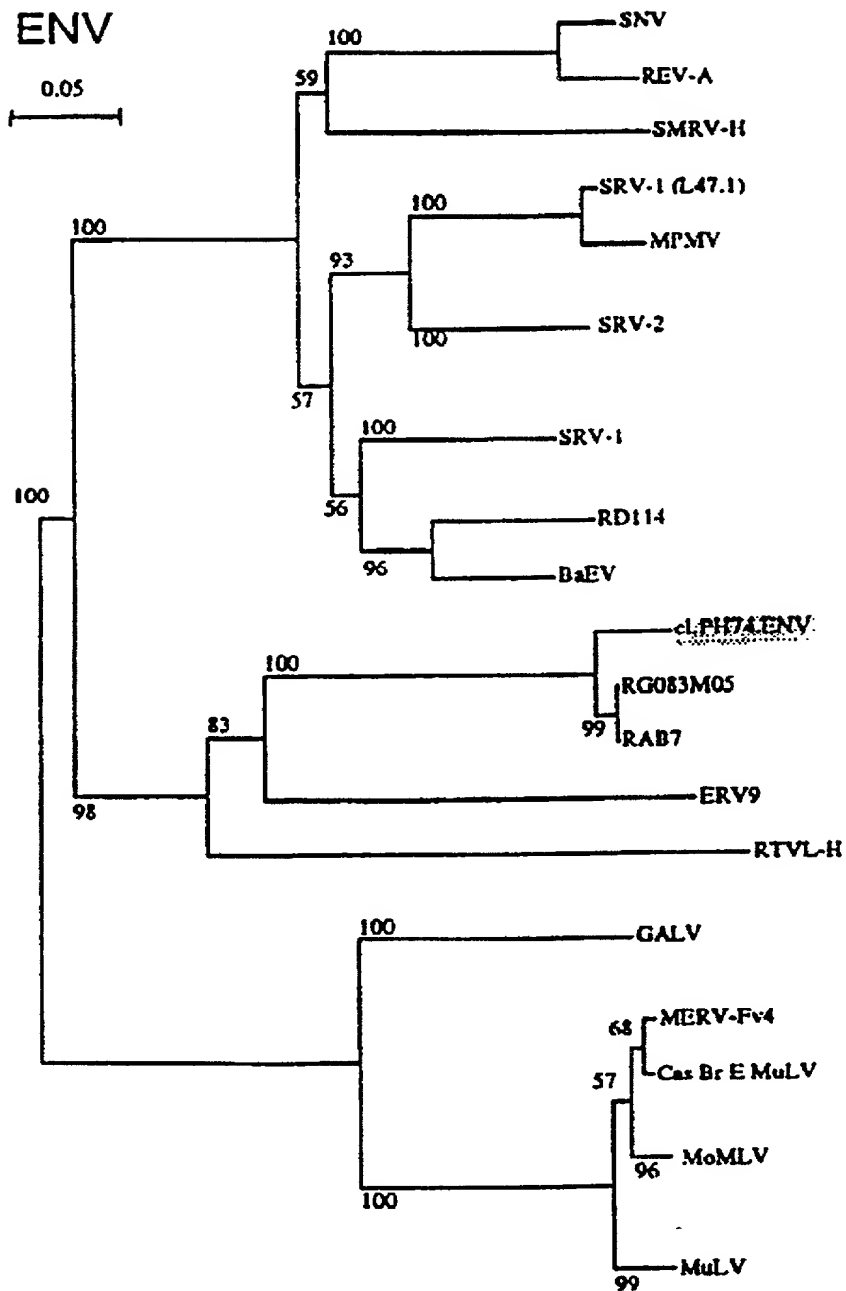


FIG 4C



## Consensus

[illegible]

12

[illegible]

**A12**

347  
332  
358  
359  
74  
3  
ATG

513

[illegible]

•

## FIG5B

5-RG-28000-28872  
3-RG-37500-38314  
5-6A2.1-600  
5-PH74.1-530  
5-24.4.1-486

Consensus

CCGACGACCTGCGGCTGACTCCCATCCCTCTGATCTCTGAGGTGTCCGCTGTGCTCTGATCCAGCGGCGCCATTGCGGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  
CCGACGACCTGCGGCTGACTCCCATCCCTCTGATCTCTGAGGTGTCCGCTGTGCTCTGATCCAGCGGCGCCATTGCGGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  
CCGACGACCTGCGGCTGACTCCCATCCCTCTGATCTCTGAGGTGTCCGCTGTGCTCTGATCCAGCGGCGCCATTGCGGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  
CCGACGACCTGCGGCTGACTCCCATCCCTCTGATCTCTGAGGTGTCCGCTGTGCTCTGATCCAGCGGCGCCATTGCGGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  
CCGACGACCTGCGGCTGACTCCCATCCCTCTGATCTCTGAGGTGTCCGCTGTGCTCTGATCCAGCGGCGCCATTGCGGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC  
-----U5-----  
CCGACGACCTGCGGCTGACTCCCATCCCTCTGATCTCTGAGGTGTCCGCTGTGCTCTGATCCAGCGGCGCCATTGCGGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGC

585  
572  
312  
241  
198  
520

5-RG-28000-28872  
3-RG-37500-38314  
5-6A2.1-600  
5-PH74.1-530  
5-24.4.1-486

Consensus

ACGGCTAAGTGCCCTGGGTTTGTCTTAATTGAGCTGAACACTAGTCACTGGGTTTCCATGGTTCCTCTCTGTGACCCACGGCTTCTAATAGAGCTATATACACTTACCCATGCGCCCAAGATT  
ATGGCTAAGTGCCCTGGGTTTGTCTTAATTGAGCTGAACACTAGTCACTGGGTTTCCATGGTTCCTCTCTGTGACCCACGGCTTCTAATAGAGCTATATACACTTACCCATGCGCCCAAGATT  
ACGGCTAAGTGCCCTGGGTTTGTCTTAATTGAGCTGAACACTAGTCACTGGGTTTCCATGGTTCCTCTCTGTGACCCACGGCTTCTAATAGAGCTATATACACTTACCCATGCGCCCAAGATT  
ACGGCTAAGTGCCCTGGGTTTGTCTTAATTGAGCTGAACACTAGTCACTGGGTTTCCATGGTTCCTCTCTGTGACCCACGGCTTCTAATAGAGCTATATACACTTACCCATGCGCCCAAGATT  
ACGGCTAAGTGCCCTGGGTTTGTCTTAATTGAGCTGAACACTAGTCACTGGGTTTCCATGGTTCCTCTCTGTGACCCACGGCTTCTAATAGAGCTATATACACTTACCCATGCGCCCAAGATT  
-----U5-----  
AYGGCTAAGTGCCCTGGGTTTGTCTTAATTGAGCTGAACACTAGTCACTGGGTTTCCATGGTTCCTCTCTGTGACCCACGGCTTCTAATAGAGCTATATACACTTACCCATGCGCCCAAGATT

705  
692  
432  
361  
318  
640

5-RG-28000-28872  
3-RG-37500-38314  
5-6A2.1-600  
5-PH74.1-530  
5-24.4.1-486

Consensus

CCATTCTTGGATTCCTCTTARGSCACGACACTTCCASGTCTCAGAGAAACAGAGCTTGCACCACTTTGGAGCGGCTGTACTCCATCTTGGAGCTGTGTTTCCACCACTCTTGGAGCTCTTG  
CCATTCTTGGATTCCTCTTARGSCACGACACTTCCASGTCTCAGAGAAACAGAGCTTGCACCACTTTGGAGCGGCTGTACTCCATCTTGGAGCTGTGTTTCCACCACTCTTGGAGCTCTTG  
CCATTCTTGGATTCCTCTTARGSCACGACACTTCCASGTCTCAGAGAAACAGAGCTTGCACCACTTTGGAGCGGCTGTACTCCATCTTGGAGCTGTGTTTCCACCACTCTTGGAGCTCTTG  
CCATTCTTGGATTCCTCTTARGSCACGACACTTCCASGTCTCAGAGAAACAGAGCTTGCACCACTTTGGAGCGGCTGTACTCCATCTTGGAGCTGTGTTTCCACCACTCTTGGAGCTCTTG  
CCATTCTTGGATTCCTCTTARGSCACGACACTTCCASGTCTCAGAGAAACAGAGCTTGCACCACTTTGGAGCGGCTGTACTCCATCTTGGAGCTGTGTTTCCACCACTCTTGGAGCTCTTG  
-----U5-----  
CCATTCTTGGATTCCTCTTARGSCACGACACTTCCASGTCTCAGAGAAACAGAGCTTGCACCACTTTGGAGCGGCTGTACTCCATCTTGGAGCTGTGTTTCCACCACTCTTGGAGCTCTTG

824  
766  
551  
481  
437  
760

5-RG-28000-28872  
3-RG-37500-38314  
5-6A2.1-600  
5-PH74.1-530  
5-24.4.1-486

Consensus

TGAGCAAGGACCCCGGTAAACATTTTGGCAACACGAGCGGACATCCA  
TGAGCAAGGACCCCGGTAAACATTTTGGCAACACGAGGTGCAATGCAATGGG  
TGAGCAAGGACCCCGGTAAACATTTTGGCAACACGAGCGGACATCCA  
TGAGCAAGGACCCCGGTAAACATTTTGGCAACACGAGCGGACATCCA  
TGAGCAAGGACCCCGGTAAACATTTTGGCAACACGAGCGGACATCCA  
-----PBS----->  
TGAGCAAGGACCCCGGTAAACATTTTGGCAACACGAGGTGCAATGCAATGGG

873  
815  
600  
530  
486  
783

## ORF1: ENV (538 AA) FIG 6

```

<--- L --->--- SU
MGLPYHIFLCSVLSPCFTLTAPPPCRMTSSSPHPEFLWRMQRPGNIDAPSYRSLSKGTP 60
A FT V S YQ C

TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVWCWYFTQTGMSDGGGV 120

QDQAREKHVKEVISQLTG VHGTS SPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180
R

QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNVEITHTSNLTVCVKF 240
L

SNTTYTTNSQCIRWVTPPTQIVCLPSGIFVFCGTSAYRCLNGSSESMCFLSFLVPPMAIY 300
T

--->--- TM
TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTTSTQFYKLSQELNGDMER 360

VADSLVTLQDQLNSLA AVVLQNRRLDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVEE 420
R S K

IPDRIQRIAEELRNTGPWGLLSRWMPWILPFLGPLAAIILLLLFGPCIFDLLVNFVSSRI 480
R R Q N

EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSAGSS 538
--->

```

## ORF2 (52AA)

MEPKMQSKTKIYRRPLDRPVSPRSDVNDIKGTPPEEISAAQPLLRPNSAGSS-

Alignment ORF2 and Rex\_PLV-L

```

ORF2          KIY-RRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRP
++Y LD P SP ++ P S QPLLRP
Rex PTLV-L (B53482) RLYNTLSLDSPPSPPKELPA-----PSRFSPPOPLLRP

```

## ORF3 (48AA)

MLMTSKAPLLRKSQ LHNLYYAPIQQEAVRAVVGQPPQHLGFPVEMGD

Alignment ORF3 and Tat\_SIV-AGM

```

ORF3          MTSKAPLLRKSQ LHNLYYAPIQQEAVRAVVGQPPQ
+T AP R+ ++ +L AP+Q +++ G+ Q
Tat SIV-AGM(p05913) VTYHAPRTRRKKIRSLNLAPLQHSISTKWGRDQG

```



**DECLARATION AND POWER OF ATTORNEY  
UNDER 35 USC §371(c)(4) FOR  
PCT APPLICATION FOR UNITED STATES PATENT**

As a below named inventor, I hereby declare that:  
my residence, post office address and citizenship are as stated below under my name;

I verily believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought, namely the invention entitled: Endogenetic retroviral sequences, associated with autoimmune diseases or with pregnancy disorders

described and claimed in international application number \_\_\_\_\_  
filed \_\_\_\_\_.

I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations §1.56. Under Title 35, U.S. Code §119, the priority benefits of the following foreign application(s) filed within one year prior to my international application are hereby claimed:

French patent application No 97 08815 filed July 7, 1997

The following application(s) for patent or inventor's certificate on this invention were filed in countries foreign to the United States of America either (a) more than one year prior to my international application, or (b) before the filing date of the above-named foreign priority application(s):

I hereby appoint the following as my attorneys of record with full power of substitution and revocation to prosecute this application and to transact all business in the Patent Office:

James A. Oliff, Reg. No. 27,075; William P. Berridge, Reg. No. 30,024;  
Kirk M. Hudson, Reg. No. 27,562; Thomas J. Pardini, Reg. No. 30,411; and  
Edward P. Walker, Reg. No. 31,450.

**ALL CORRESPONDENCE IN CONNECTION WITH THIS APPLICATION SHOULD BE SENT TO OLIFF & BERRIDGE, P.O. BOX 19928, ALEXANDRIA, VIRGINIA 22320, TELEPHONE (703) 836-6400.**

I hereby declare that I have reviewed and understand the contents of this Declaration, and that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Typewritten Full Name L00  
of Sole or First Inventor Frédéric BESEME  
Given Name Middle Initial Family Name  
Inventor's Signature Frédéric Beseme  
Date of Signature November 8, 1999  
Residence Villefontaine FRANCE FRX  
City State or Province Country  
Citizenship French  
Post Office Address 39 rue de la Noyera  
(Insert complete mailing address, including country) 38090 Villefontaine, FRANCE

Note to Inventor: Please sign name on line 2 exactly as it appears in line 1 and insert the actual date of signing on line 3.

IF THERE IS MORE THAN ONE INVENTOR USE PAGE 2 AND PLACE AN "X" HERE ☒

(Discard this page in a sole inventor application)

Typewritten Full Name of Joint Inventor 2-00 Jean-Luc BLOND  
Given Name Middle Initial Family Name  
Inventor's Signature Jean-Luc  
Date of Signature November, 8, 1999  
Residence LYON FRANCE FRX  
City State or Province Country  
Citizenship French  
Post Office Address 75 bis rue des Acqueducs  
(Insert complete mailing address, including country) 69005 Lyon, FRANCE

Typewritten Full Name of Joint Inventor 3-00 Olivier BOUTON  
Given Name Middle Initial Family Name  
Inventor's Signature Olivier  
Date of Signature November, 8, 1999  
Residence Francheville FRANCE FRX  
City State or Province Country  
Citizenship French  
Post Office Address 48 Avenue du Châter  
(Insert complete mailing address, including country) 69340 Francheville, FRANCE

Typewritten Full Name of Joint Inventor 4-00 Bernard MANDRAND  
Given Name Middle Initial Family Name  
Inventor's Signature Bernard  
Date of Signature November, 8, 1999  
Residence Villeurbanne FRANCE FRX  
City State or Province Country  
Citizenship French  
Post Office Address 21 rue de la Doua  
(Insert complete mailing address, including country) 69100 Villeurbanne, FRANCE

Typewritten Full Name of Joint Inventor 5-00 François MALLET  
Given Name Middle Initial Family Name  
Inventor's Signature François  
Date of Signature November 8, 1999  
Residence Villeurbanne FRANCE FRX  
City State or Province Country  
Citizenship French  
Post Office Address 84 rue Anatole France  
(Insert complete mailing address, including country) 69100 Villeurbanne, FRANCE

Typewritten Full Name of Joint Inventor 6-00 Hervé PERRON  
Given Name Middle Initial Family Name  
Inventor's Signature Hervé  
Date of Signature November 8, 1999  
Residence Lyon FRANCE FRX  
City State or Province Country  
Citizenship French  
Post Office Address 15 rue de Boyer  
(Insert complete mailing address, including country) 69005 Lyon, FRANCE

Note to Inventor: Please sign name on line 2 exactly as it appears in line 1 and insert the actual date of signing on line 3.

This form may be executed only when attached to the first page of the Declaration and Power of Attorney of the application to which it pertains.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## 5 (i) APPLICANT:

(A) NAME: BIO MERIEUX

(B) STREET: CHEMIN DE L'ORME

(C) CITY: MARCY L'ETOILE

(E) COUNTRY: FRANCE

10 (F) POSTAL CODE: 69280

(ii) TITLE OF INVENTION: NUCLEIC MATERIAL OF THE  
ENDOGENOUS RETROVIRAL GENOMIC TYPE, ASSOCIATED WITH AN  
AUTOIMMUNE DISEASE AND/OR WITH PREGNANCY DISORDERS; USE  
15 AS MARKER

(iii) NUMBER OF SEQUENCES: 35

## 20 (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0,  
Version #1.30 (EPO)

25

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1321 base pairs

30 (B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA (as DNA)

35

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CAACAATCGG GATATAAACC CAGGCATTCT AGCTGGCAAC AGCAGCCCCC CTTTGGGTCC 60  
CTTCCCTTTG TATGGGAGCT GTTTTCATGC TATTTCACTC TATTAAATCT TGCAACTGCA 120  
CTCTTCTGGT CCATGTTTCT TACGGCTCGA GCTGAGCTTT TGCTCACCGT CCACCACTGC 180  
TGTTTGCCAC CACCGCAGAC CTGCCGCTGA CTCCCATCCC TCTGGATCCT GCAGGGTGTC 240  
CGCTGTGCTC CTGATCCAGC GAAGCGCCCA TTGCCGCTCC CAATTGGGCT AAAGGCTTGC 300  
CATTGTTTCT GCACGGCTAA GTGCCTGGGT TTGTTCTAAT TGAGCTGAAC ACTAGTCACT 360  
GGGTTCCATG GTTCTCTTCT GTGACCCACG GCTTCTAATA GAACTATAAC ACTTACCACA 420  
TGGCCCAAGA TTCCATTCTT TGAATCCGT GAGGCCAAGA ACTCCAGGTC AGAGAATACG 480  
AAGCTTGCCA CCATCTTGGA AGCGGCCTGC TACCATCTTG GAAGTGGTTC ACCACCATCT 540  
TGGGAGCTCT GTGAGCAAGG ACCCCCCGGT AACATTTTGG CAACCAAGAA CGGACATCCA 600  
AAGTGATGGG AAACGTTCCC CGCAAGACAA AAACGCCCCCT AAGACGTATT CTGGAAAATT 660  
GGGAACAATT TGACCCCTCAG AACTAAGAA AGAAACGACT TATATTCTTC TGCAGTGCCG 720  
CCTGGCACTC CTGAGGGAAG TATAAATTAT AACACCATCT TACAGCTAGA CCTCTTTTGT 780  
AGAAAAGGCA AATGGAGTGA AGTGCCATAA GTACAACTT TCTTTTCATT AAGAGACAAC 840  
TCACAATTAT GTAAAAAGTC TGATTTATGC CCTACAGGAA GCCTTCAGAG TCTACCTCCC 900  
TATCCCAGCA TCCCCGACTC CTTCCCCACT TAATAAGGAC CCCCCTTCAA CCCAAATGGT 960  
CCAAAAGGAG ATAGACAAAA GGTAAACAG TGAACCAAAG AGTGCCAATA TTCCCCAATT 1020

0946024-12169

ATGACCCCTC CAAGCAGTGG GAGGAAGAGA ATTCGGCCCA GCCAGAGTGC ATGTGCCTTT 1080  
TTCTCTCCCA GACTTAAAGC AAATAAAAAC AGACTTAGGT AAATTCTCAG ATAACCCTGA 1140  
TGGCTATATT GGTGTTTTAC AAGGGTTAGG ACAATTCTTT GATCTGACAT GGAGAGATAT 1200  
ATATGTCACT GCTAAATCAG AACTAACCC CAAATGAGAG AAGTGCCACC ATAACTGCAG 1260  
CCTGAGAGTT TGGCGATCTC TGGTATCTCA GTCAGGTCAA TGATAGGATG ACAACAGAGG 1320  
A 1321

(2) INFORMATION FOR SEQ ID NO: 2:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2938 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
10 (ii) MOLECULE TYPE: mRNA (as DNA)  
(iii) HYPOTHETICAL: NO  
15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CAACGACGGA CATCCAAAGT GATGGGAAAC GTTCCCCGCA AGACAAAAAC GCCCCTAAGA 60  
CGTATTCTGG AGAATTGGGA CCAATTTGAC CCTCAGACAC TAAGAAAGAA ACGACTTATA 120  
TTCTTCTGCA GTGCCGCCTG GCACTCCTGA GGGAAGTATA AATTATAACA CCATCTTACA 180  
GCTAGACTTC TTTTGTAGAA AAGGCAAATG GAGTGAAGTG CCATAAGTAC AAACCTTCTT 240

TTCATTAAGA GACAACTCAC AATTATGTAA AAAGTGTGAT TTATGCCCTA CAGGAAGCCT 300  
 TCAGAGTCTA CCTCCCTATC CCAGCATCCC CGACTCCTTC CCCAACTAAT AAGGACCCCC 360  
 CTTCAACCCA AATGGTCCAA AAGGAGATAG ACAAAGGGT AAACAGTGAA CCAAAGAGTG 420  
 CCAATATTCC CCAATTATGA CCCCTCCCAA GCAGTGGGAG GAAGAGATTC GGCCAGCCA 480  
 GAGTGCATGT GCTTTTCTT CTCCCAGACT TAAAGCAAAT AAAACAGAC TTAGGTAAAT 540  
 TCTCAGATAA TCCTGATGGC TATATTGATG TTTTACAAGG GTTAGGACAA TTCTTTGATC 600  
 TGACATGGAG AGATATAATG TCACTGCTAA ATCAGACACT AACCCCAAAT GAGAGAAGTG 660  
 CCACCATAAC TGCAGCCTGA GAGTTGGCG ATCTCTGGTA TCTCAGTCAG GTCAATGATA 720  
 GGATGACAAC AGAGGAAAGA GATGATCCCC ACAGCCAGCA AGCAGTTCCC AGTCTASACC 780  
 CTCATTGGGG ACACAGAAAT CAGTAACATG GGAGATTGGT GCTGCAGACA TTTGCTAACT 840  
 TGTGTGCTAC AAGGACTAAG GAAAACTACG AAGAAAATCT ACGAATTACT CAATGATGTC 900  
 CACCATAACA CAGGGGAAGG GAAGAAAATC CTACTGCCTT TCTGGAGAGA CTAAGGGAGG 960  
 CATTGAGGAA GCGTGCCTCT CTGTCACCTG ACTCTTCTGA AGGCCAACTA ATCTTAAAGC 1020  
 GTAAGTTTAT CACTCAGTCA GCTGCAGACA TTAGAAAAA CTTCAAAAGT CTGCCGTAGG 1080  
 CCCGGAGCAA AACTTAGAAA CCCTATTGAA CTTGGCAACY TCGGTTTTTT ATAATAGAGA 1140  
 TCAGGAGGAG CAGGCGGAAC AGGACAAACG GGATTAAAAA AAAGGCCACC GCTTTAGTCA 1200  
 TGACCCTCAG GCAAGTGGAC TTTGGAGGCT CTGGAAAAGG GAAAAGCTGG GCAAATTGAA 1260  
 TGCCTAATAG GGCTTGCTTC CAGTGC GGTC TACAAGGACA CTTTAAAAA GATTGTCCAA 1320

66927-12094150

GTAGAAGTAA GCCGCCCCCTT CGTCCATGCC CTTATTTCA AGGGAATCAC TGAAGGCCC 1380  
ACTGCCCCAG GGGACAAAGG TCTTTTGAGT CAGAAGCCAC TAACCAGATG ATCCAGCAGC 1440  
AGGACTGAGG GTGCCTGGGG CAAGCGCCAT CCCATGCCAT CACCCTCACA GAGCCCTGGG 1500  
TATGCTTGAC CATTGAGGGC CAGGAAGGTT GTCTCCTGGA CACTGGTGCG GTCTTCTTAG 1560  
TCTTACTCTT CTGTCCCGGA CAACTGTCCT CCAGATCTGT CACTATCTGA GGGGGTCCTA 1620  
AGACGGGCAG TCACTAGATA CTTCTCCCAG CCACTAAGTT ATGACTGGGG AGCTTTATTC 1680  
TTTTCACATG CTTTTCTAAT TATGCTTGAA AGCCCCACTA CTTGTTAGG GAGAGACATT 1740  
CTAGCAAAAG CAGGGGCCAT TATACACCTG AACATAGGAG AAGGAACACC CGTTTGTGT 1800  
CCCCTGCTTG AGGAAGGAAT TAATCCTGAA GTCTGGGCAA CAGAAGGACA ATATGGACGA 1860  
GCAAAGAATG CCCGTCTGT TCAAGTTAA CTAAAGGATT CCACTTCCTT TCCCTACCAA 1920  
AGGCAGTACC CCCTCAGACC CAAGGCCCAA CAAGGATTCC AAAAGATTGT TAAGGACTTA 1980  
AAAGCCCAAG GCTTAGTAA ACCATGCATA ACTCCCTGCA GTAATCCGT AGTGGATTGA 2040  
GGAGGCACAG AAACCCAGTG GACAGTGGAG GGTAGTGCA AGATCTCAGG ATTATCAATG 2100  
GAGGCCGTTG TCCTTTTATA CCCAGCTGTA CCTAGCCCTT ATACTGTGCT TTCCCAAATA 2160  
CCAGAGGAAG CAGAGTGGTT TACACTCCTG GACCTTAAGG ATGCCTTCTT CTGCATCCCT 2220  
GTACATCCTG ACTCTCAATT CTTGTTTGCC TTTGAAGATA CTTCAAACCC AACATCTCAA 2280  
CTCACCTGGA CTGTTTTACC CCAAGGGTTC AGGATAGCC CCCATCTATT TGGCCAGGCA 2340  
TTAGCCCAAG ACTTGAGCCA ATCCTCATAC CTGGACACTT GTCCTTCGGT AGGTGGATGA 2400

TTTACTTTTG GCCGCCCATTT CAGAAACCTT GTGCCATCAA GCCACCCAAG CGCTCTTCAA 2460  
TTTCCTCGCT ACCTGTGGCT ACATGGTTTC CAAACCAAAG GCTCAACTCT GCTCACAGCA 2520  
GGTTACTTAG GGCTAAAATT ATCCAAAGGC ACCAGGGCCC TCAGTGAGGA ACACATCCAG 2580  
CCTATACTGG CTTATCCTCA TCCCAAACC CTAAAGCAAC TAAGGGGATT CCTTGGCGTA 2640  
ATAGGTTTCT GCCGAAAATG GATTCCCAGG TTTGGCGAAA TAGCCAGGTC ATTAAATACA 2700  
CTAATTAAGG AACTCAGAA AGCCAATACC CATTTAGTAA GATGGACAAC TGAAGTAGAA 2760  
GTGGCTTTCC AGGCCCTAAC CCAAGCCCCA GTGTTAAGTT TGCCAACAGG GCAAGACTTT 2820  
TCTTCATATG TCACAGAAAA AACAGGAATA GCTCTAGGAG TCCTTACACA GATCCGAGGG 2880  
ATGAGCTTGC AACCTGTGGC GTACCTGACT AAGGAAATTG ATGTAGTGGC AAAGGGTT 2938

(2) INFORMATION FOR SEQ ID NO: 3:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1422 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: mRNA (as DNA)
- (iii) HYPOTHETICAL: NO
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TCAGGGATAG CCCCATCTA TTTGGCCAGG CATTAGCCCA AGACTTGAGT CAGTTATCAT 60  
ACCTGGACAC TCTTGTCTT CAGTATGTGG ATGATTTACT TTTAGCTGCC TGTTTCAGAA 120



CCTTGTGCCA TCAAGCCACC CAAGCACTCT TAAATTTCTT CGCCACCTGT GGCTACAAGG 180  
TTTCCAAAGA GAAGCTCAGC TCTGCTCACA GCAGGTAAA TACTTAGGAC TAAGATTATC 240  
CAAAGGCACC AAGGCCCTCA GTGAGGAATG TATCCAGCCT ATACTGGCTT ATCCTCATCT 300  
CAAAACCCTA AAGCAACTAA GAGAGTTCCT TGGCATAACA GGCTTCTGCC GAATATGGAT 360  
TCCCCAGGTA TGGCAAATA GCCAGGCCAT TATATACAGT AATTAAGGAA ACTCAGAAAG 420  
CCAATACCCA TTTAATAAGA TGGATACCTG AAGCCAAAGT GGCTTTCCAG GCCCCTAAAG 480  
AAGGCCTTAA ACCCAAGTCC CAGTGTTAAG CTTGCCAAG GGGCAAGACT TTTCTTTATA 540  
CATCACAGAA AAAACAGAA ACAGCTCTGG GAGTCCTTAC ACAGGTCCAA GGGACGAGCT 600  
TGCAACCCAT GGCATACCTG AGTAAGGAAA CTGATGTAGT GGCAAAGGCT TGGCTTCATT 660  
GTTTATGGGT AGTGGTGGCA GTAGCAGTTG TAGTATCTGA AGCAGTTAAA ATAATACAGG 720  
GGAGAGATCT TACTGTGTGG ACATCTCATG AGGTGAACAG CATACTCACT GCTAAAGGAG 780  
ACTTGTGGCT GTCAGACAAC CGTTTACTTA AATATCAGGC TCTATTACTT GAAAGGCCAG 840  
TGCTGCAACT GTGCACTTGT GCAACTCTTA ACCCAGTCNC ATTTCTTCCA GACAATGAAG 900  
ATAGAATATA ACTGTCAACA AATAATTTCT CAAACCTATG CCACTCGAGG GCACCTTCTA 960  
GAAGTTCCTT TGAATGATCC TGACCTTCAA CTTGTATACT GATGGAAGTT CCTTTGTAGA 1020  
AAAAGGACTT CAAAAGCGGG GTATGCAGTG GTCAGTGATA ATGGAATATT TGAAAGTATC 1080  
CCCTCACTCC AGGAACTAGT GCTTAGCTGG CAGAACTAAT AGCCTTCATT GGGGCACTAG 1140  
AATTAGGAGA AGGAAAAGG GTAAATATAT ATACAGACTC TGAGTATGCT CACCTAGTCN 1200

TCCATGCCCA TGAGGCAATA TGCAGAGAAA GGAATTCCT AACTCCGAG GGAACACCTA 1260  
TCACACATCA GGAAGCCATT AGGAGATTAT TACTGGCAGT ACAGAAACCT AAAGAGGTGG 1320  
AAGTCTTACA CTGCTGGGGT CATCAGAAAG GAAAGAAAAG GGAATAGAA GGAATTGCC 1380  
AAGCAGATAT TGAAGCAAAA AGAGCTGCAA GGCAGGACCC TC 1422

(2) INFORMATION FOR SEQ ID NO: 4:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2006 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- 10 (ii) MOLECULE TYPE: mRNA (as DNA)

- (iii) HYPOTHETICAL: NO

- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATGCAGTGGT CAGTGATAAT GGAATACTTG AAAGTAATCC CCTCACTCCA GGAAGTAGTG 60  
CTCAGCTAGC AGAACTAATA GCCCTCACTT GGGCACTAGA ATTAGGAGAA GAAAAAAGGG 120  
CAAATATATA TACAGACTCT AAATATGCTT ACCTAGTCCT CCATGCCCAT GCAGCAATAT 180  
GGAAAGAAAG GGAATTCCTA ACTTCTGAGA GAACACCTAT CAAACATCAG GAAGCCATTA 240  
GGAAATTATT ATTGGCTGTA CAGAAACCTA AAGAGGTGGC AGTCTTACAC TGCCGGGGTC 300  
ATCANAAAGG AAAGGAAAGG GAAAATACTT TTGCETGCAA CTATCCAATG GAAATTACTT 360

AAAACCCTTC ATCAAACCTT TCACTTAGGC ATCGATAGCA CCCATCAAAT GGCCAAATCA 420  
TTATTTACTG GACCAGGCCT TTTCAAACT ATCAAGCAA TATTCAGGGC CTGTGAATTG 480  
TGCCAAAAAA ATAATCCCCT GCCTCATCGC CAAGCTCCTT CAGGAAAACA AAAACAGGC 540  
CATTACCCTG AAAAAACTG GCAACTGATT TTACCCACAA GCCCAAACCT CAGGGATTTT 600  
AGTATCTACT AGTCTGGGTA AATACTTTCA CGGGTTGGGC AAAGGCCTTC CCCTGTAGGA 660  
CAGAAAAGGC CCAAGAGGTA ATAAAGGCAC TAGTTCATGA AATAATTCCC AGATTCGGAC 720  
TTCCCGAGG CTTACAGAGT GACAATAGCC CTGCTTTCCA GGCCACAGTA ACCCAGGGAG 780  
TATCCAGGC GTTAGGTATA CGATATCACT TACTGCGC CTGAAGGCCA CAGTCCTCAG 840  
GGAAGGTCGA GAAATGAAT GAAATACTCA AAGGACATCT AAAAAAGCAA ACCCAGGAAA 900  
CCCACCTCAC ATGGCCTGCT CTGTTGCCTA TAGCCTTAA AAGAATCTGC AACTTTCCCC 960  
AAAAAGCAGG ACTTAGCCCA TACGAAATGC TGTATGGAAG GCCCTTCATA ACCAATGACC 1020  
TTGTGCTTGA CCCAAGACAG CCAACTTAGT TGCAGACATC ACCTCCTTAG CCAAATATCA 1080  
ACAAGTTCTT AAAACATTAC AAGGAACCTA TCCCTGAGAA GAGGGAAAAG AACTATTCCA 1140  
CCCTGTGAC ATGGTATTAG TCAAGTCCCT TCTCTCTAAT TCCCCATCCC TAGATACATC 1200  
CTGGGAAGGA CCCTACCCAG TCATTTTATT TACCCCAACT GCGGTTAAAG TGGCTGGAGT 1260  
GGTCTTGGAT ACATCACACT TGAGTCAAAT CCTGGATACT GCCAAAGGAA CCTGAAAATC 1320  
CAGGAGACAA CGCTAGCTAT TCCTGTGAAC CTCTAGAGGA TTTGCGCCTG CTCTTCAAAC 1380  
AACAACCAGG AGGAAAGTAA CTAAAATCAT AAATCCCCCA TGGCCCTCCC TTATCATATT 1440

TTTCTCTTTA CTGTTCTTTT ACCCTCTTTC ACTCTCACTG CACCCCTCC ATGCCGCTGT 1500  
 ATGACCAGTA GCTCCCCTTA CCAAGAGTTT CTATGGAGAA TGCAGCGTCC CGGAAATATT 1560  
 GATGCCCCAT CGTATAGGAG TCTTTCTAAG GGAACCCCCA CCTTCACTGC CCACACCCAT 1620  
 ATGCCCCGCA ACTGCTATCA CTCTGCCACT CTTTGCATGC ATGCAAATAC TCATTATTGG 1680  
 ACAGGAAAAA TGATTAATCC TAGTTGTCTT GGAGGACTTG GAGTCACTGT CTGTTGGAAT 1740  
 TACTTCACCC AACTGGTAT GTCTGATGGG GGTGGAGTTC AAGATCAGGC AAGAGAAAAA 1800  
 CATGTAAAG AAGTAATCTC CCAACTCACC CGGTACATG GCACCTCTAG CCCTACAAAG 1860  
 GACTAGATCT CTCAAAATA CATGAAACCC TCCGTACCCA TACTCGCCTG GTAAGCCTAT 1920  
 TTAATACCAC CCTCACTGGG CTCCATGAGG TCTCGGCCCA AAACCCTACT AACTGTTGGA 1980  
 TATGCCTCCC CCTGAACTTC AAGCCA 2006

(2) INFORMATION FOR SEQ ID NO: 5:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1948 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ACTGCACTCT TCTGGTCCAT GTTCTTACG GCTCGAGCTG AGCTTTTGCT CACCGTCCAC 60

CACTGCTGTT TGCCACCACC GCANACCTGC CGCTGACTCC CATCCCTCTG GATCCTGCAG 120  
GGTGTCCGCT GTGCTCCTGA TCCAGCGAGG CGCCCATTGC CGCTCCCAAT TGGGCTAAAG 180  
GCTTGCCATT GTNCCTGCAC GGCTAAGTGC CTGGGTTTGT TCTAATTGAG CTGAACACTA 240  
NTCACTGGGT TCCATGGTTC TCTTCTGTGA CCCACGGCTT CTAATAGAAC TATAACACTT 300  
ACCACATGGC CCAAGATTCC ATTCCTTGGA ATCCGTGAGG GCAAGAACTC CAGGTCAGAG 360  
AATACGAGGC TTGCCACCAT CTGGAAGCG GCCTGCTACC ATCTTGAAG TGGTTCACCA 420  
CCATCTTGGG AGCTCTGTGA GCAAGGACCC CCCGGTAACA TTTTGGCAAC CACGAACGGA 480  
CATCCAAAGT GATACATCCT GGAAGGACC CTACCCAGTC ATTTTATCTA CCCCAACTGC 540  
GGTTAAAGTG GCTGGAGTGG AGTCTTGAT ACATCACACT TGAGTCAAAT CCTGGATACT 600  
GCCAAAGGAA CCTGAAAATC CAGGAGACAA CGCTAGCTAT TCCTGTGAAC CTCTAGAGGA 660  
TTTGCGCCTG CTCTTCAAAC AACACCAGG AGGAAAGTAA CTAAATCAT AAATCCCAT 720  
GGCCCTCCCT TATCATATTT TTCTCTTTAC TGTGTTTCA CCCTCTTTCA CTCTCACTGC 780  
ACCCCTCCA TGCCGCTGTA TGACCAGTAG CTCCCCTTAC CAAGAGTTTC TATGGAGAAT 840  
GCAGCGTCCC GGAAATATTG ATGCCCCATC GTATAGGAGT CTTTGTAAGG GAACCCCCAC 900  
CTTCACTGCC CACACCCATA TGCCCCGCAA CTGCTATCAC TCTGCCACTC TTTGCATGCA 960  
TGCAAATACT CATTATTGGA CAGGAAAAAT GATTAATCCT AGTTGTCCTG GAGGACTTGG 1020  
AGTCACTGTC TGTGGAATT ACTTCACCCA AACTGGTATG TCTGATGGGG GTGGAGTTCA 1080  
AGATCAGGCA AGAGAAAAAC ATGTAAAAGA AGTAATCTCC CAACTACCC GGGTACATGG 1140

CACCTCTAGC CCCTACAAAG GACTAGATCT CTCAAAATA CATGAAACCC TCCGTACCCA 1200  
TACTCGCCTG GTAAGCCTAT TTAATACCAC CCTCACTGGG CTCCATGAGG TCTCGGCCCA 1260  
AAACCCTACT AACTGTTGGA TATGCCTCCC CCTGAACTTC AGGCCATATG TTTCAATCCC 1320  
TGTACCTGAA CAATGGAACA ACTTCAGCAC AGAAATAAAC ACCACTTCCG TTTTAGTAGG 1380  
ACCTCTTGTT TCCAATCTGG AAATAACCCA TACCTCAAAC CTCACCTGTG TAAAATTTAG 1440  
CAATACTAGA TACACAACCA ACTCCCAATG CATCAGGTGG GTAACCTCCTC CCACACAAAT 1500  
AGTCTGCCTA CCCTCAGGAA TATTTTTTGT CTGTGGTACC TCAGCCTATC GTTGTGTTGAA 1560  
TGGCTCTTCA GAATCTATGT GCTTCCTCTC ATTCTTAGTG CCCCCTATGG CCATCTACAC 1620  
TGAACAAGAT TTATACAGTT ATGTCATATC TAAGCCCCGC AACAAAAGAG TACCCATTCT 1680  
TCCTTTTGTT ATAGGAGCAG GAGTGCTAGG TGCACTAGGT ACTGGCATTG GCGGTATCAC 1740  
AACCTCTACT CAGTTCTACT ACAAATATC TCAAGAACTA AATGGGGACA TGGAACGGGT 1800  
CGCCGACTCC CTGGTCACCT TGCAAGATCA ACTTAACTCC CTAGCAGCAG TAGTCCTTCA 1860  
AAATCGAAGA GCTTTAGACT TGCTAACCGC TGAAGAGGGG GGAACCTGTT TATTTTtagg 1920  
GGAAGAATGC TGTATTATG TTAATCAA 1948

(2) INFORMATION FOR SEQ ID NO: 6:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1136 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCATGGCCAT CTACACTGAA CAAGATTTAT ACAGTTATGT CATATCTAAG CCCC GCAACA 60  
AAAGAGTACC CATTCTTCCT TTTGTTATAG GAGCAGGAGT GCTAGGTGCA CTAGGTA CTG 120  
GCATTGGCGG TATCACAACC TCTACTCAGT TCTACTACAA ACTATCTCAA GAACTAAATG 180  
GGGACATGGA ACGGGTCGCC GACTCCCTGG TCACCTTGCA AGATCAACTT AACTCCCTAG 240  
CAGCAGTAGT CCTTCAAAAT CGAAGAGCTT TAGACTCGCT AACCGCTGAA AGAGGGGGAA 300  
CCTGTTTATT TTTAGGGGAA GAATGCTGTT ATTATGTTAA TCAATCCGGA ATCGTCACTG 360  
AGAAAGTTAA AGAAATTCGA GATCGAATAC AACGTAGAGC AGAAGAGCTT CGAAACACTG 420  
GACCCTGGGG CCTCTCAGC CAATGGATGC CCTGGATTCT CCCCTTCTTA GGACCTCTAG 480  
CAGCTATAAT ATTGCTACTC CTCTTTGGAC CCTGTATCTT TAACCTCCTT GTTAACCTTG 540  
TCTCTCCAG AATCGAAGCT GTAAACTAC AAATGGAGCC CAAGATGCAG TCCAAGACTA 600  
AGATCTACCG CAGACCCCTG GACCGGCCTG CTAGCCACG ATCTGATGTT AATGACATCA 660  
AAGGCACCCC TCCTGAGGAA ATCTCAGCTG CACAACCTCT ACTACGCCCC AATTCAGCAG 720  
GAAGCAGTTA GAGCGGTCGT CGGCCAACCT CCCCACAGC ACTTAGGTTT TCCTGTTGAG 780  
ATGGGGGACT GAGAGACAGG ACTAGCTGGA TTTCTAGGC TGACTAAGAA TCCCTAAGCC 840  
TAGCTGGGAA GGTGACCACA TCCACCTTTA AACACGGGGC TTGCAACTTA GTTCACACCT 900  
GACCAATCAG AGAGCTCACT AAAATGCTAA TTAGGCAAAG ACAGGAGGTA AAGAAATAGC 960  
CAATCATCTA TTGCATGAGA GCACAGCAGG AGGGACAATG ATCGGGATAT AAACCCAAGT 1020  
CTTCGAGCCG GCAACGGCAA CCCCCTTTGG GTCCCTCCC TTTGTATGGG AGCTCTGTTT 1080  
TCATGCTATT TCACTCTATT AAATCTTGCA GCTGCGAAAA AAAAAAAAAA AAAAAA 1136

5

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 2782 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

65 92 129 166 203 240 277 314 351 388 425 462 499 536 573 610 647 684 721 758 795 832 869 906 943 980 1017 1054 1091 1128 1165 1202 1239 1276 1313 1350 1387 1424 1461 1498 1535 1572 1609 1646 1683 1720 1757 1794 1831 1868 1905 1942 1979 2016 2053 2090 2127 2164 2201 2238 2275 2312 2349 2386 2423 2460 2497 2534 2571 2608 2645 2682 2719 2756 2793 2830 2867 2904 2941 2978 3015 3052 3089 3126 3163 3200 3237 3274 3311 3348 3385 3422 3459 3496 3533 3570 3607 3644 3681 3718 3755 3792 3829 3866 3903 3940 3977 4014 4051 4088 4125 4162 4199 4236 4273 4310 4347 4384 4421 4458 4495 4532 4569 4606 4643 4680 4717 4754 4791 4828 4865 4902 4939 4976 5013 5050 5087 5124 5161 5198 5235 5272 5309 5346 5383 5420 5457 5494 5531 5568 5605 5642 5679 5716 5753 5790 5827 5864 5901 5938 5975 6012 6049 6086 6123 6160 6197 6234 6271 6308 6345 6382 6419 6456 6493 6530 6567 6604 6641 6678 6715 6752 6789 6826 6863 6900 6937 6974 7011 7048 7085 7122 7159 7196 7233 7270 7307 7344 7381 7418 7455 7492 7529 7566 7603 7640 7677 7714 7751 7788 7825 7862 7899 7936 7973 8010 8047 8084 8121 8158 8195 8232 8269 8306 8343 8380 8417 8454 8491 8528 8565 8602 8639 8676 8713 8750 8787 8824 8861 8898 8935 8972 9009 9046 9083 9120 9157 9194 9231 9268 9305 9342 9379 9416 9453 9490 9527 9564 9601 9638 9675 9712 9749 9786 9823 9860 9897 9934 9971 10008 10045 10082 10119 10156 10193 10230 10267 10304 10341 10378 10415 10452 10489 10526 10563 10600 10637 10674 10711 10748 10785 10822 10859 10896 10933 10970 11007 11044 11081 11118 11155 11192 11229 11266 11303 11340 11377 11414 11451 11488 11525 11562 11599 11636 11673 11710 11747 11784 11821 11858 11895 11932 11969 12006 12043 12080 12117 12154 12191 12228 12265 12302 12339 12376 12413 12450 12487 12524 12561 12598 12635 12672 12709 12746 12783 12820 12857 12894 12931 12968 13005 13042 13079 13116 13153 13190 13227 13264 13301 13338 13375 13412 13449 13486 13523 13560 13597 13634 13671 13708 13745 13782 13819 13856 13893 13930 13967 14004 14041 14078 14115 14152 14189 14226 14263 14300 14337 14374 14411 14448 14485 14522 14559 14596 14633 14670 14707 14744 14781 14818 14855 14892 14929 14966 15003 15040 15077 15114 15151 15188 15225 15262 15299 15336 15373 15410 15447 15484 15521 15558 15595 15632 15669 15706 15743 15780 15817 15854 15891 15928 15965 16002 16039 16076 16113 16150 16187 16224 16261 16298 16335 16372 16409 16446 16483 16520 16557 16594 16631 16668 16705 16742 16779 16816 16853 16890 16927 16964 17001 17038 17075 17112 17149 17186 17223 17260 17297 17334 17371 17408 17445 17482 17519 17556 17593 17630 17667 17704 17741 17778 17815 17852 17889 17926 17963 18000 18037 18074 18111 18148 18185 18222 18259 18296 18333 18370 18407 18444 18481 18518 18555 18592 18629 18666 18703 18740 18777 18814 18851 18888 18925 18962 19000 19036 19073 19110 19147 19184 19221 19258 19295 19332 19369 19406 19443 19480 19517 19554 19591 19628 19665 19702 19739 19776 19813 19850 19887 19924 19961 20000 20036 20073 20110 20147 20184 20221 20258 20295 20332 20369 20406 20443 20480 20517 20554 20591 20628 20665 20702 20739 20776 20813 20850 20887 20924 20961 21000 21036 21073 21110 21147 21184 21221 21258 21295 21332 21369 21406 21443 21480 21517 21554 21591 21628 21665 21702 21739 21776 21813 21850 21887 21924 21961 22000 22036 22073 22110 22147 22184 22221 22258 22295 22332 22369 22406 22443 22480 22517 22554 22591 22628 22665 22702 22739 22776 22813 22850 22887 22924 22961 23000 23036 23073 23110 23147 23184 23221 23258 23295 23332 23369 23406 23443 23480 23517 23554 23591 23628 23665 23702 23739 23776 23813 23850 23887 23924 23961 24000 24036 24073 24110 24147 24184 24221 24258 24295 24332 24369 24406 24443 24480 24517 24554 24591 24628 24665 24702 24739 24776 24813 24850 24887 24924 24961 25000 25036 25073 25110 25147 25184 25221 25258 25295 25332 25369 25406 25443 25480 25517 25554 25591 25628 25665 25702 25739 25776 25813 25850 25887 25924 25961 26000 26036 26073 26110 26147 26184 26221 26258 26295 26332 26369 26406 26443 26480 26517 26554 26591 26628 26665 26702 26739 26776 26813 26850 26887 26924 26961 27000 27036 27073 27110 27147 27184 27221 27258 27295 27332 27369 27406 27443 27480 27517 27554 27591 27628 27665 27702 27739 27776 27813 27850 27887 27924 27961 28000 28036 28073 28110 28147 28184 28221 28258 28295 28332 28369 28406 28443 28480 28517 28554 28591 28628 28665 28702 28739 28776 28813 28850 28887 28924 28961 29000 29036 29073 29110 29147 29184 29221 29258 29295 29332 29369 29406 29443 29480 29517 29554 29591 29628 29665 29702 29739 29776 29813 29850 29887 29924 29961 30000 30036 30073 30110 30147 30184 30221 30258 30295 30332 30369 30406 30443 30480 30517 30554 30591 30628 30665 30702 30739 30776 30813 30850 30887 30924 30961 31000 31036 31073 31110 31147 31184 31221 31258 31295 31332 31369 31406 31443 31480 31517 31554 31591 31628 31665 31702 31739 31776 31813 31850 31887 31924 31961 32000 32036 32073 32110 32147 32184 32221 32258 32295 32332 32369 32406 32443 32480 32517 32554 32591 32628 32665 32702 32739 32776 32813 32850 32887 32924 32961 33000 33036 33073 33110 33147 33184 33221 33258 33295 33332 33369 33406 33443 33480 33517 33554 33591 33628 33665 33702 33739 33776 33813 33850 33887 33924 33961 34000 34036 34073 34110 34147 34184 34221 34258 34295 34332 34369 34406 34443 34480 34517 34554 34591 34628 34665 34702 34739 34776 34813 34850 34887 34924 34961 35000 35036 35073 35110 35147 35184 35221 35258 35295 35332 35369 35406 35443 35480 35517 35554 35591 35628 35665 35702 35739 35776 35813 35850 35887 35924 35961 36000 36036 36073 36110 36147 36184 36221 36258 36295 36332 36369 36406 36443 36480 36517 36554 36591 36628 36665 36702 36739 36776 36813 36850 36887 36924 36961 37000 37036 37073 37110 37147 37184 37221 37258 37295 37332 37369 37406 37443 37480 37517 37554 37591 37628 37665 37702 37739 37776 37813 37850 37887 37924 37961 38000 38036 38073 38110 38147 38184 38221 38258 38295 38332 38369 38406 38443 38480 38517 38554 38591 38628 38665 38702 38739 38776 38813 38850 38887 38924 38961 39000 39036 39073 39110 39147 39184 39221 39258 39295 39332 39369 39406 39443 39480 39517 39554 39591 39628 39665 39702 39739 39776 39813 39850 39887 39924 39961 40000 40036 40073 40110 40147 40184 40221 40258 40295 40332 40369 40406 40443 40480 40517 40554 40591 40628 40665 40702 40739 40776 40813 40850 40887 40924 40961 41000 41036 41073 41110 41147 41184 41221 41258 41295 41332 41369 41406 41443 41480 41517 41554 41591 41628 41665 41702 41739 41776 41813 41850 41887 41924 41961 42000 42036 42073 42110 42147 42184 42221 42258 42295 42332 42369 42406 42443 42480 42517 42554 42591 42628 42665 42702 42739 42776 42813 42850 42887 42924 42961 43000 43036 43073 43110 43147 43184 43221 43258 43295 43332 43369 43406 43443 43480 43517 43554 43591 43628 43665 43702 43739 43776 43813 43850 43887 43924 43961 44000 44036 44073 44110 44147 44184 44221 44258 44295 44332 44369 44406 44443 44480 44517 44554 44591 44628 44665 44702 44739 44776 44813 44850 44887 44924 44961 45000 45036 45073 45110 45147 45184 45221 45258 45295 45332 45369 45406 45443 45480 45517 45554 45591 45628 45665 45702 45739 45776 45813 45850 45887 45924 45961 46000 46036 46073 46110 46147 46184 46221 46258 46295 46332 46369 46406 46443 46480 46517 46554 46591 46628 46665 46702 46739 46776 46813 46850 46887 46924 46961 47000 47036 47073 47110 47147 47184 47221 47258 47295 47332 47369 47406 47443 47480 47517 47554 47591 47628 47665 47702 47739 47776 47813 47850 47887 47924 47961 48000 48036 48073 48110 48147 48184 48221 48258 48295 48332 48369 48406 48443 48480 48517 48554 48591 48628 48665 48702 48739 48776 48813 48850 48887 48924 48961 49000 49036 49073 49110 49147 49184 49221 49258 49295 49332 49369 49406 49443 49480 49517 49554 49591 49628 49665 49702 49739 49776 49813 49850 49887 49924 49961 50000 50036 50073 50110 50147 50184 50221 50258 50295 50332 50369 50406 50443 50480 50517 50554 50591 50628 50665 50702 50739 50776 50813 50850 50887 50924 50961 51000 51036 51073 51110 51147 51184 51221 51258 51295 51332 51369 51406 51443 51480 51517 51554 51591 51628 51665 51702 51739 51776 51813 51850 51887 51924 51961 52000 52036 52073 52110 52147 52184 52221 52258 52295 52332 52369 52406 52443 52480 52517 52554 52591 52628 52665 52702 52739 52776 52813 52850 52887 52924 52961 53000 53036 53073 53110 53147 53184 53221 53258 53295 53332 53369 53406 53443 53480 53517 53554 53591 53628 53665 53702 53739 53776 53813 53850 53887 53924 53961 54000 54036 54073 54110 54147 54184 54221 54258 54295 54332 54369 54406 54443 54480 54517 54554 54591 54628 54665 54702 54739 54776 54813 54850 54887 54924 54961 55000 55036 55073 55110 55147 55184 55221 55258 55295 55332 55369 55406 55443 55480 55517 55554 55591 55628 55665 55702 55739 55776 55813 55850 55887 55924 55961 56000 56036 56073 56110 56147 56184 56221 56258 56295 56332 56369 56406 56443 56480 56517 56554 56591 56628 56665 56702 56739 56776 56813 56850 56887 56924 56961 57000 57036 57073 57110 57147 57184 57221 57258 57295 57332 57369 57406 57443 57480 57517 57554 57591 57628 57665 57702 57739 57776 57813 57850 57887 57924 57961 58000 58036 58073 58110 58147 58184 58221 58258 58295 58332 58369 58406 58443 58480 58517 58554 58591 58628 58665 58702 58739 58776 58813 58850 58887 58924 58961 59000 59036 59073 59110 59147 59184 59221 59258 59295 59332 59369 59406 59443 59480 59517 59554 59591 59628 59665 59702 59739 59776 59813 59850 59887 59924 59961 60000 60036 60073 60110 60147 60184 60221 60258 60295 60332 60369 60406 60443 60480 60517 60554 60591 60628 60665 60702 60739 60776 60813 60850 60887 60924 60961 61000 61036 61073 61110 61147 61184 61221 61258 61295 61332 61369 61406 61443 61480 61517 61554 61591 61628 61665 61702 61739 61776 61813 61850 61887 61924 61961 62000 62036 62073 62110 62147 62184 62221 62258 62295 62332 62369 62406 62443 62480 62517 62554 62591 62628 62665 62702 62739 62776 62813 62850 62887 62924 62961 63000 63036 63073 63110 63147 63184 63221 63258 63295 63332 63369 63406 63443 63480 63517 63554 63591 63628 63665 63702 63739 63776 63813 63850 63887 63924 63961 64000 64036 64073 64110 64147 64184 64221 64258 64295 64332 64369 64406 64443 64480 64517 64554 64591 64628 64665 64702 64739 64776 64813 64850 64887 64924 64961 65000 65036 65073 65110 65147 65184 65221 65258 65295 65332 65369 65406 65443 65480 65517 65554 65591 65628 65665 65702 65739 65776 65813 65850 65887 65924 65961 66000 66036 66073 66110 66147 66184 66221 66258 66295 66332 66369 66406 66443 66480 66517 66554 66591 66628 66665 66702 66739 66776 66813 66850 66887 66924 66961 67000 67036 67073 67110 67147 67184 67221 67258 67295 67332 67369 67406 67443 67480 67517 67554 67591 67628 67665 67702 67739 67776 67813 67850 67887 67924 67961 68000 68036 68073 68110 68147 68184 68221 68258 68295 68332 68369 68406 68443 68480 68517 68554 68591 68628 68665 68702 68739 68776 68813 68850 68887 68924 68961 69000 69036 69073 69110 69147 69184 69221 69258 69295 69332 69369 69406 69443 69480 69517 69554 69591 69628 69665 69702 69739 69776 69813 69850 69887 69924 69961 70000 70036 70073 70110 70147 70184 70221 70258 70295 70332 70369 70406 70443 70480 70517 70554 70591 70628 70665 70702 70739 70776 70813 70850 70887 70924 70961 71000 71036 71073 71110 71147 71184 71221 71258 71295 71332 71369 71406 71443 71480 71517 71554 71591 71628 71665 71702 71739 71776 71813 71850 71887 71924 71961 72000 72036 72073 72110 72147 72184 72221 72258 72295 72332 72369 72406 72443 72480 72517 72554 72591 72628 72665 72702 72739 72776 72813 72850 72887 72924 72961 73000 73036 73073 73110 73147 73184 73221 73258 73295 73332 73369 73406 73443 73480 73517 73554 73591 73628 73665 73702 73739 73776 73813 73850 73887 73924 73961 74000 74036 74073 74110 74147 74184 74221 74258 74295 74332 74369 74406 74443 74480 74517 74554 74591 74628 74665 74702 74739 74776 74813 74850 74887 74924 74961 75000 75036 75073 75110 75147 75184 75221 75258 75295 75332 75369 75406 75443 75480 75517 75554 75591 75628 75665 75702 75739 75776 75813 75850 75887 75924 75961 76000 7



TCCATTCCCTT GGAATCCGTG AGGCCAACGA ACTCCAGGTC AGAGAATACG AAGCTTGCCA 420

CCATCTTGGA AGCGGCCTGC TACCATCTTG GAAGTGGTTC ACCACCATCT TGGGAGCTCT 480

GTGAGCAAGG ACCCCCCGGT GACATTTTGG CGACCACCAA CGGACATCCC AAGTGATACA 540

TCCTGGGAAG GACCCTACCC AGTCATTTTA TCTACCCCAA CTGCGGTAA AGTGGCTGGA 600

GTGGAGTCTT GGATACATCA CACTTGAGTC AAATCCTGGA TACTGCCAAA GGAACCTGAA 660

AATCCAGGAG ACAACGCTAG CTATTCCTGT GAACCTCTAG AGGATTGCG CCTGCTCTTC 720

AAACAACAAC CAGGAGGAAA GTAACATAAA TCATAAATCC CCATGGGCCT CCCTTATCAT 780

ATTTTCTCT GTAGTGTCTT TTCACCCTGT TCACTCTCA CTGCACCCCC TCCATGCCGC 840

TGTATGACCA GTAGCTCCCC TCACCAGAG TTTCTATGGA GAATGCAGCG TCCCGGAAAT 900

ATTGATGCCC CATCGTATAG GAGTCTTTCT AAGGGAACCC CCACCTTAC TGGCCACACC 960

CATATGCCCC GCAACTGCTA TCACTCTGCC ACTCTTTGCA TGCATGCAAA TACTCATTAT 1020

TGGACAGGAA AAATGATTAA TCCTAGTTGT CCTGGAGGAC TTGGAGTCAC TGTCTGTTGG 1080

ACTTACTTCA CCCAACTGG TATGTCTGAT GGGGGTGGAG TTCAAGATCA GGCAAGAGAA 1140

AAACATGTAA AAGAAGTAAT CTCCCAACTC ACCGGGGTAC ATGGCACCTC TAGCCCCTAC 1200

AAAGGACTAG ATCTCTCAA ACTACATGAA ACCCTCCGTA CCCATACTCG CCTGGTAAGC 1260

CTATTTAATA CCACCCTCAC TGGGCTCCAT GAGGTCTCGG CCCAAAACCC TACTAACTGT 1320

TGGATATGCC TCCCCCTGAA CTTCAGGCCA TATGTTTCAA TCCCTGTACC TGAACAATGG 1380

AACAACCTCA GCACAGAAAT AAACACCACT TCCGTTTTAG TAGGACCTCT TGTTTCCAAT 1440

GTGGAAATAA CCCATACCTC AAACCTCACC TGTGTAAAT TTAGCAATAC TACATACACA 1500  
ACCAACTCCC AATGCATCAG GTGGGTAACCT CCTCCACAC AAATAGTCTG CCTACCCTCA 1560  
GGAATATTTT TTGTCTGTGG TACCTCAGCC TATCGTTGTT TGAATGGCTC TTCAGAATCT 1620  
ATGTGCTTCC TCTCATTCTT AGTGCCCCCT ATGACCATCT ACACTGAACA AGATTATAC 1680  
AGTTATGTCA TATCTAAGCC CCGCAACAAA AGAGTACCCA TTCTTCCTTT TGTATAGGA 1740  
GCAGGAGTGC TAGGTGCACT AGGTACTGGC ATTGGCGGTA TCACAACCTC TACTCAGTTC 1800  
TACTACAAAC TATCTCAAGA ACTAAATGGG GACATGGAAC GGGTCGCCGA CTCCTGGTTC 1860  
ACCTTGCAAG ATCAACTTAA CTCCCTAGCA GCAGTAGTCC TTCGAAATCG AAGAGCTTTA 1920  
GACTTGCTAA CCGCTGAGAG AGGGGGAACC TGTTTATTTT TAGGGGAAGA ATGCTGTTAT 1980  
TATGTTAATC AATCCGGAAT CGTCACTGAG AAAGTTGAAG AAATCCAGA TCGAATACAA 2040  
CGTATAGCAG AGGAGCTTCG AAACACTGGA CCCTGGGGCC TCCTCAGCCG ATGGATGCCC 2100  
TGGATTCTCC CCTTCTTAGG ACCTCTAGCA GCTATAATAT TGCTACTCCT CTTTGGACCC 2160  
TGTATCTTTG ACCTCCTTGT TAACTTTGTC TCTCCAGAA TCGAAGCTGT GAAACTACAA 2220  
ATGGAGCCCA AGATGCAGTC CAAGACTAAG ATCTACCGCA GACCCCTGGA CCGGCCTGCT 2280  
AGCCCACGAT CTGATGTTAA TGACATCAAA GGCACCCCTC CTGAGGAAAT CTCAGCTGCA 2340  
CAACCTCTAC TACGCCCCAA TTCAGCAGGA AGCAGTTAGA GCGGTGGTCG GCCAACCTCC 2400  
CCAACAGCAC TTAGGTTTTT CTGTTGAGAT GGGGGACTGA GAGACAGGAC TAGCTGGATT 2460  
TCCTAGGCTG ACTAAGAATC CTTAAGCCTA GGTGGGAAGG TGACCACATC CACCTTTAAA 2520

CACGGGGCTT GCAACTTAGC TCACACCTGA CCAATCAGAG AGCTCACTAA AATGCTAATT 2580  
AGGCAAAGAC AGGAGGTAAA GAAATAGCCA ATCATTATT GCCTGAGAGC ACAGCAGGAG 2640  
GGACAATGAT CGGGATATAA ACCCAAGTTT TCGAGCCGGC AACGGCAACC CCCTTTGGGT 2700  
CCCCTCCCTT TGTATGGGAG CTCTGTTTTT ATGCTATTTC ACTCTATTAA ATCTTGCAAC 2760  
TGCAAAAAAA AAAAAAAAAA AA 2782

(2) INFORMATION FOR SEQ ID NO: 8:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 666 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
10 (ii) MOLECULE TYPE: mRNA (as DNA)  
(iii) HYPOTHETICAL: NO  
15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TGTCCGCTGT GCTCCTGATC CAGCGAGGCG CCCATTGCCG CTCCCAATTG GGCTAAAGGC 60  
TTGCCATTGT TCCTGCACGG CTAAGTGCCT GGGTTTGTTT TAATTGAGCT GAACACTANT 120  
CACTGGGTTC CATGGTTCTC TTCTGTGACC CACGGCTTCT AATATACTA TAACACTTAC 180  
CACATGGCCC AAGATTCCAT TCCTTGGAAT CCGTGAGGCC AAGAACTCCA GGTCAGAGAA 240  
TACGAGGCTT GCCACCATCT TGGAAGCGGC CTGCTACCAT CTTGGAAGTG GTTCACCACC 300

ATCTTGGGAG CTCTGTGAGC AAGGACCCCC CGGTAACATT TTGGCAACCA CGAACGGACA 360  
 TCCAAAGTGA ATCGAAGCTG TAAAACTACA AATGGAGCCC AAGATGCAGT CCAAGACTAA 420  
 GATCTACCGC AGACCCCTGG ACCGGCCTGC TAGCCCACGA TCTGATGTTA ATGACATCAA 480  
 AGGCACCCCT CCTGAGGAAA TCTCAGCTGC ACAACCTCTA CTACGCCCCA ATTGAGCAGG 540  
 AAGCAGTTAG AGCGGTCGTC GGCCAACTC CCCAACAGCA CTTAGGTTTT CCTGTTGAGA 600  
 TGGGGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCTAAGCCT 660  
 AGCTGG 666

(2) INFORMATION FOR SEQ ID NO: 9:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3372 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: mRNA (as DNA)
- (iii) HYPOTHETICAL: NO
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GACTTCCCAA ATACCAGAGG AAGCAGAGTG GTTACAGTC CTGGACCTTC AGGATGCCTT 60  
 CTTCTGCATC CCTGTACATC CTGACTCTCA ATTCTGTTT GCCTTTGAAG ATACTTCAAA 120  
 CCCAGCATCT CAACTCACCT GGACTATTTT ACCCCAAGGG TTCAGGGATA GTCCCCATCT 180  
 ATTTGGCCAG GCATTAGCCC AAGACTTGAG CCAATCCTCA TACCTGGACA CTTGTCCTTC 240

GGTAGGTGGA TGATTACTT TTGGCCGCCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC 300  
AAGCGCTCTT CAATTTCCTC GCTACCTGTG GCTACATGGT TTCCAAACCA AAGGCTCAAC 360  
TCTGCTCACA GCAGGTACT TAGGGCTAAA ATTATCCAAA GGCACCAGGG CCCTCAGTGA 420  
GGAACACATC CAGCCTATAC TGGCTTATCC TCATCCCAA ACCCTAAAGC AACTAAGGGG 480  
ATTCCTTGGC GTAATAGGTT TCTGCCGAAA ATGGATTCCC AGGTATGGCG AAATAGCCAG 540  
GTCATTAAAT AACTAATTA AGGAACTCA GAAAGCCAAT ACCCATTTAG TAAGATGGAC 600  
AACTGAAGTA GAAGTGGCTT TCCAGGCCCT AACCAGGCC CCAGTGTTAA GTTTGCCAAC 660  
AGGGCAGAC TTTTGTTCAT ATGTCACAGA AAAAACAGGA ATAGCTCTAG GAGTCCTTAC 720  
ACAGATCCGA GGGATGAGCT TGCAACCTGT GGCACACCTG ACTAAGGAAA TTGATGTAGT 780  
GGCAAAGGGT TGACCTCATT GTTTACGGGT AGTGGTGGCA GTAGCAGTCT TAGTATCTGA 840  
AGCAGTTAAA ATAATACAGG GAAGAGATCT TACTGTGTGG ACATCTCATG ATGTGAATGG 900  
CATACTCACT GCTAAAGGAG ACTTGTGGCT GTCAGACRAC TGTTTACTTA AATGTCAGGC 960  
TCTATTACTT GAAGGGCCAG TGCTGCGACT GTGCACTTGT GCAACTCTTA ACCCAGCCAC 1020  
ATTTCTTCCA GACAATGAAG AAAAGATAAA ACATAACTGT CAACAAGTAA TTTCTCAAAC 1080  
CTATGCCACT CGAGGGGACC TTTTAGAGGT TCCTTTGACT GATCCCGACC TCAACTTGTA 1140  
TACTGATGGA AGTTCCTTTG TAGAAAAAGG ACTTCGAAAA GTGGGGTATG CAGTGGTCAG 1200  
TGATAATGGA AACTTGAAA GTAATCCCCT CACTCCAGGA ACTAGTGCTC AGCTAGCAGA 1260  
ACTAATAGCC CTCACTTGGG CACTAGAATT AGGAGAAGAA AAAAGGGCAA ATATAATACA 1320

GACTCTAAAT ATGCTTACCT AGTCCTCCAT GCCCATGCAG CAATATGGAA AGAAAGGGAA 1380  
TTCCTAACTT CTGAGAGAAC ACCTATCAAA CATCAGGAAG CCATTAGGAA ATTATTATTG 1440  
GCTGTACAGA AACCTAGAGA GGTGGCAGTC TTACACTGCC GGGGTCATCA CAAAGGAAAG 1500  
GAAAGGGAAA TACAAGAGAA CTGCCAAGCA TATATTGAAG CCAAAAGAGC TGCAAGGCAG 1560  
GACCCTCCAT TAGAAATGCT TATTAACTT CCCTTAGTAT AGGGTAATCC CTTCCGGGAA 1620  
ACCAAGCCCC AGTACTCAGC AGGAGAAACA GAATGGGGAA CCTCACGAGG CAGTTTTCTC 1680  
CCCTCGGGAC GGTTAGCCAC TGAAGAAGGG AAAATACTTT TGCCTGCAAC TATCCAATGG 1740  
AAATTACTTA AAACCCTTCA TCAAACCTTT CACTTAGGCA TCGATAGCAC CCATCAGATG 1800  
GCCAATCAT TATTTACTGG ACCAGGCCTT TTCAAACTA TCAAGCAGAT AGTCAGGGCC 1860  
TGTGAAGTGT GCCAGAGAAA TAATCCCCTG CCTTATCGCC AAGCTCCTTC AGGAGAACAA 1920  
AGAACAGGCC ATTACCCTGG AGAAGACTGG CAACTGATTT TACCCACAAG CCCAACCTC 1980  
AGGGATTTC A GTATCTACTA GTCTGGGTAG ATACTTTCAC GGGTTGGGCA GAGGCCTTCC 2040  
CCTGTAGGAC AGAAAAGGCC CAAGAGGTAA TAAAGGCACT AGTTCATGAA ATAATTCCCA 2100  
GATTCCGACT TCCCCGAGGC TTACAGAGTG ACAATAGCCC TGCTTCCAG GCCACAGTAA 2160  
CCCAGGGAGT ATCCCAGGCG TTAGGTATAC GATATCACTT AACTGCGCC TGAAGGCCAC 2220  
AGTCCTCAGG GAAGGTCGAG AAAATGAATG AAACACTCAA AGGACATCTA AAAAGCAAA 2280  
CCCAGGAAAC CCACCTCACA TGGCCTGTTC TGTGCCTAT AGCCTTAAAA AGAATCTGCA 2340  
ACTTTCCCA AAAAGCAGGA CTTAGCCCAT ACGAAATGCT GTATGGAAGG CCCTTCATAA 2400

CCAATGACCT TGTGCTTGAC CCAAGACAGC CAACTTAGTT GCAGACATCA CCTCCTTAGC 2460

CRAATATCAA CAAGTTCTTA AAACATTACA AGGACCTAT CCCTGAGAAG AGGAAAAGAA 2520

TATTCCACCC AAGTGACATG GTATTAGTCA AGTCCCTTCC CTCTAATTCC CCATCCCTAG 2580

ATACATCCTG GGAAGGACCC TACCCAGTCA TTTTATCTAC CCCAACTGCG GTTAAAGTGG 2640

CTGGACTGGA GTCTTGATA CATCACACTT GAGTCAAATC CTGGATACTG CCAAAGGAAC 2700

CTGAAAATCC AGGAGACAAC GCTAGCTATT CCTGTGAACC TCTAGAGGAT TTGCGCCTGC 2760

TCTTCAAACA ACAACCAGGA GGAAAAATCG AAGCTGTAAA ACTACAAATG GAGCCCAAGA 2820

TGCAGTCCAA GACTAAGATC TACCGCAGAC CCCTGGACCG GCCTGTTAGC CCACGATCTG 2880

ATGTTAATGA CATCAAAGGC ACCCCTCCTG AGGAAATCTC AGCTGCACAA CCTCTACTAC 2940

GCCCCAATTC AGCAGGAAGC AGTTAGAGCG GTCGTCGGCC AACCTCCCCA ACAGCACTTA 3000

GGTTTTCTCG TTGAGATGGG GGAAGTGAAC ACAGGACTAG CTGGATTTCCT TAGGCTGATT 3060

AAGAATCCCT AAGCCTAGCT GGAAGGTGA CCACATCCAC CTTTAAACAC GGGGCTTGCA 3120

ACTTAGCTCA CACCTGACCA ATCAGAGAGC TCACTAAAAT GCTAATTAGG CAAAGACAGG 3180

AGCTAAAGAA ATAGCCAATC ATTTATTGCC TGAGAGCACA GCAGGAGGGA CAATGATCGG 3240

GATATAAACC CAAGTTTTCG AGCCGGCAAC GGCAACCCCC TTTGGGTCCC CTCCCTTTGT 3300

ATGGGAGCTC TGTTTTCATG CTATTTCACT CTATTAAATC TTGCAACTGC AAAAAAAAAA 3360

AAAAAAAAAA AA 3372

(2) INFORMATION FOR SEQ ID NO: 10:

- 5 (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2372 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

REPLACEMENT SHEET (RULE 26)

69429460

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

ACTGCACTCT TCTGGTCCAT GTTCTTACG GCTCGAGCTG AGCTTTTGCT CACCGTCCAC 60  
CACTGCTGTT TGCCACCACC GCAGACCTGC CGCTGACTCC CATCCCTCTG GATCCTGCAG 120  
GGTGTCCGCT GTGCTCCTGA TCCAGCGAGG CGCCCATTCG CGCTCCCAAT TGGGCTAAAG 180  
GCTTGCCATT GTTCCTGCAC GGCTAAGTGC CTGGGTTTGT TCTAATTGAG CTGAACACTA 240  
ATCACTGGGT TCCATGGTTC TCTTCTGTGA CCCACGGCTT CTAATAGAAC TATAACACTT 300  
ACCACATGGC CCAAGATTCC ATTCCTTGGA ATCCGTGAGG CCAAGAACTC CAGGTCAGAG 360  
AATACGAGGC TTGCCACCAT CTTGGAAGCG GCCTGCTACC GTCTTGGAAG TGGTTCACCA 420  
CCATCTTGGG AGCTCTGTGA GCAAGGACCC CCCGTAACA TTTTGGCAAC CAACGACGGA 480  
CATCCAAAGT GATGGGAAAC GTTCCCCGCA AGACAAAAC GCCCCTAAGA CGTATTCTGG 540  
AGAATTGGGA CCAATTTGAC CCTCAGACAC TAAGAAAGAA ACGACTTATA TTCTTCTGCA 600  
GTGCCGCCTG GCACTCCTGA GGGAAGTATA AATTATAACA CCATCTTACA GCTAGACCTC 660



TTTTGTAGAA AAGGCAAATG GAGTGAAGTG CCATAAGTAC AAACTTTCTT TTCATTAAGA 720

GACAACTCAC AATTATGTAA AAAGTGTGAT TTATGCCCTA CAGGAAGCCT TCAGAGTCTA 780

CCTCCCTATC CCAGCATCCC CGACTCCTTC CCCAACTAAT AAGGACCCCC CTTCAACCCA 840

AATGGTCCAA AAGGAGATAG ACAAAGGGT AAACAGTGAA CCAAAGAGTG CCAATATTCC 900

CCAATTATGA CCCCTCCAAG CAGTGGGAGG AAGAGAATTC GGCCAGCCA GAGTGCATGT 960

GCCTTTTTCT CTCCCAGACT TAAAGCAAAT AAAACAGAC TTAGGTAAAT TCTCAGATAA 1020

CCCTGATGGC TATATTGATG TTTTACAAGG GTTAGGACAA TTCTTTGATC TGACATGGAG 1080

AGATATAATG TCACTGCTAA ATCAGACACT AACCCCAAAT GAGAGAAGTG CCACCATAAC 1140

TGCAGCCTGA GGGTTTGGCG TCTCTGGTAT CTCAGTCAGG TCAATGGATA NGGATGACAA 1200

CAGAAGGAAA GANAATGATT CCCACAGGC CAGCAGGCAG TTCCCAGTCT AGACCCTCAT 1260

TGGGACACAG AATCAGAACA TGGAGATTGG TGCTGCAGAC ATTTGCTAAC TTGTGTGCTA 1320

GAAGGACTAA GGAAACTAG GAAGAAGTCT ATGAATTACT CAATGATGTC CACCATAACA 1380

CAGGGAAGGG AAGAAAATCC TACTGCCTTT CTGGAGAGAC TAAGGGAGGC ATTGAGGAAG 1440

CGTGCCTCTC TGTACCTGA CTCTTCTGAA GGCCAACTAA TCTTAAAGCG TAAGTTTATC 1500

ACTCAGTCAG CTGCAGACAT TAGAAAAAC TTCAAAGTC TGCCGTAGGC CCGGAGCAA 1560

ACTTAGAAAC CCTATTGAAC TTGGCAACCT CGGTTTTTTA TAATAGAGAT CAGGAGGAGC 1620

AGGCGGAACA GGACAAACGG GATTAAAAA AAGGCCACCG CTTTAGTCAT GACCCTCAGG 1680

CAAGTGGACT TTGGAGGCTC TGGAAAACGG AAAAGCTGGG CAAATTGAAT GCCTAATAGG 1740

GCTTGCTTCC AGTGCGGTCT ACAAGGACAC TTTAAAAAAG ATTGTCCAAG TAGAAGTAAG 1800  
CCGCCCCCTTC GTCCATGCCC CTTATTTCOA GGAATCACT GGAAGGCCCA CTGCCCCAGG 1860  
GGACAAAGGT CTTTGTAGTC AGAAGCCACT AACCAGATGA TCCAGCAGCA GGACTGAGGG 1920  
TGCCTGGGGC AAGCGCCATC CCATGCCATC ACCCTCACAG AGCCCTGGGT ATGCTTGACC 1980  
ATTGAGGGCC AGGAAGGTTG TCTCCTGGAC ACTGGTGGG TCTTCTTAGT CTTACTCTTC 2040  
TGTCCCGGAC AACTGTCTC CAGATCTGTC ACTATTCTGA GGGGGTCCNT AAGACGGGCA 2100  
GTCCTAGAT ACTTTTTCCC AGCCACTAAG TTATGAACTG GGGAGCTTTA TTCTTTTCAC 2160  
ATGCTTTTCT AATTATGCTT GAAAGCCCCA CTACCTTGTT AGGGAGAGAC ATTCTAGCAA 2220  
AAGCAGGGGC CATTATACAC CTGAACATAG GAGAAGGAAC ACCCGTTTGT TGTNCCCCTG 2280  
CTTGAGGAAG GAATTAATCC TGAAGTCTGG GCAACAGAAG GACAATATGG ACGAGCCAAA 2340  
GAATGCCCGT CCTGTTCAAG TTAAACTAAA GG 2372

(2) INFORMATION FOR SEQ ID NO: 11:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7582 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: mRNA (as DNA)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CAACAATCGG GATATAAACC CAGGCATTGG AGCTGGCAAC AGCAGCCCCC CTTTGGGTCC 60  
CTTCCCTTTG TATGGGAGCT GTTTTCATGC TATTTCACTC TATTAAATCT TGCAACTGCA 120  
CTCTTCTGGT CCATGTTTCT TACGGCTCGA GCTGAGCTTT TGCTCACCCT CCACCACTGC 180  
TGTTTGCCAC CACCGCANAC CTGCCGCTGA CTCCCATCCC TCTGGATCCT GCAGGGGTGTC 240  
CGCTGTGCTC CTGATCCAGC GARGCGCCCA TTGCCGCTCC CAATTGGGCT AAAGGCTTGC 300  
CATTGTNCC TGCACGGCTAA GTGCCTGGGT TTGTTCTAAT TGAGCTGAAC ACTANTCACT 360  
GGGTTCATG GTTCTCTTCT GTGACCCACG GCTTCTAATA KAACTATAAC ACTTACCACA 420  
TGGCCCAAGA TTCCATTCCCT TGAATCCGT GAGGSCAACG AACTCCAGGT CAGAGAATAC 480  
GARGCTTGCC ACCATCTTGG AAGCGGCCTG CTACCRCTTT GGAAGTGGTT CACCACCATC 540  
TTGGGAGCTC TGTGAGCAAG GACCCCCCGG TRACATTTTG GCRACCAMSR ACGGACATCC 600  
MAAGTGATGG GAAACGTTCC CCGCAAGACA AAAACGCCCC TAAGACGTAT TCTGGARAAT 660  
TGGGAMCAAT TTGACCCTCA GACACTAAGA AAGAAACGAC TTATATTCTT CTGCAGTGCC 720  
GCCTGGCACT CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACYTCTTTTG 780  
TAGAAAAGGC AAATGGAGTG AAGTGCCATA AGTACAACT TTCTTTTCAT TAAGAGACAA 840  
CTCACAATTA TGTA AAAAGT GTGATTTATG CCCTACAGGA AGCCTTCAGA GTCTACCTCC 900  
CTATCCCAGC ATCCCCGACT CCTTCCCCAM YTAATAAGGA CCCCCCTTCA ACCCAAATGG 960  
TCCAAAAGGA GATAGACAAA AGGGTAAACA GTGAACCAA GAGTGCCAAT ATTCCCCAAT 1020  
TATGACCCCT CCCAAGCAGT GGGAGGAAGA GAATTCGGCC CAGCCAGAGT GCATGTGCTT 1080  
TTTTTCTCC CAGACTTAAA GCAAATAAAA ACAGACTTAG GTAAATTCTC AGATAAYCCT 1140  
GATGGCTATA TTGRTGTTTT ACAAGGGTTA GGACAATTCT TTGATCTGAC ATGGAGAGAT 1200  
ATATATGTCA CTGCTAAATC AGACACTAAC CCCAAATGAG AGAAGTGCCA CCATAACTGC 1260  
AGCCTGAGRG TTTGGCGATC TCTGGTATCT CAGTCAGGTC AATGGATANG GATGACAACA 1320  
GAAGGAAAGA NAATGATTCC CCACAGGCCA GCARGCAGTT CCCAGTCTAS ACCCTCATTG 1380  
GGGACACAGA AATCAGTAAC ATGGGAGATT GGTGCTGCAG ACATTGCTA ACTTGTGTGC 1440  
TASAAGGACT AAGGAAAAC ASGAAGAAAR TCTAYGAATT ACTCAATGAT GTCCACCATA 1500  
ACACAGGGGA AGGGAAGAAA ATCCTACTGC CTTTCTGGAG AGACTAAGGG AGGCATTGAG 1560  
GAAGCGTGCC TCTCTGTCAC CTGACTCTTC TGAAGGCCAA CTAATCTTAA AGCGTAAGTT 1620  
TATCACTCAG TCAGCTGCAG ACATTAGAAA AAACCTCAA AGTCTGCCGT AGGCCCGGAG 1680  
CAAACTTAG AAACCTATT GAACTGGCA ACYTGGGTTT TTTATAATAG AGATCAGGAG 1740  
GAGCAGGCGG AACAGGACAA ACGGGATTAA AAAAAAGGCC ACCGCTTTAG TCATGACCCT 1800  
CAGGCAAGTG GACTTTGGAG GCTCTGGAAA AGGGAAGGC TGGGCAAATT GAATGCCTAA 1860  
TAGGGCTTGC TTCCAGTGCG GTCTACAAGG ACACCTTTAA AAAGATTGTC CAAGTAGAAG 1920  
TAAGCCGCC CTTCTGCCAT GCCCTTATT TCAAGGGAAT CACTGGAAG CCCACTGCCC 1980  
CAGGGGACAA AGGTCTTTTG AGTCAGAAG CACTAACCAG ATGATCCAGC AGCAGGACTG 2040  
AGGGTGCCCT GGGCAAGCGC CATCCCATGC CATCACCTC ACAGAGCCCT GGGTATGCTT 2100

GACCATTGAG GGCCAGGAAG GTTGTCTCCT GGACACTGGT GCGGTCTTCT TAGTCTTACT 2160  
 CTTCTGTCCC GGACAACCTGT CCTCCAGATC TGTCACATATT CTGAGGGGGT CCNTAAGACG 2220  
 GGCAGTCACT AGATACTTTY TCCCAGCCAC TAAGTTATGA ACTGGGGAGC TTTATTCTTT 2280  
 TCACATGCTT TTCTAATTAT GCTTGAAAGC CCCACTACCT TGTTAGGGAG AGACATTCTA 2340  
 GCAAAAGCAG GGGCCATTAT ACACCTGAAC ATAGGAGAAG GAACACCCGT TTGTTGTNCC 2400  
 CCTGCTTGAG GAAGGAATTA ATCCTGAAGT CTGGGCAACA GAAGGACAAT ATGGACGAGC 2460  
 CAAAGAATGC CCGTCCTGTT CAACTTAAAC TAAAGGATTG CACTTCCTTT CCCTACCAAA 2520  
 GGCAGTACCC CCTCAGACCC AAGGCCAAC AAGGATTCCA AAAGATTGTT AAGGACTTAA 2580  
 AAGCCCAAGG CTTAGTAAAA CCATGCATAA CTCCTGCAG TAATTCCGTA GTGGATTGAG 2640  
 GAGGCACAGA AAGCCAGTGG ACAGTGGAGG GTTAGTGCAA GATCTCAGGA TTATCAATGG 2700  
 AGGCCGTTGT CTTTTTATAC CCAGCTGTAC CTAGCCCTTA TACTGTGMYT TCCCAAATAC 2760  
 CAGAGGAAGC AGAGTGGTTT ACASTCCTGG ACCTTMAGGA TGCCTTCTTC TGCATCCCTG 2820  
 TACATCCTGA CTCTCAATTC TTGTTTGCTT TTGAAGATAC TTCAAACCCA RCATCTCAAC 2880  
 TCACCTGGAC TRTTTTACCC CAAGGGTTCA GGGATAGYCC CCATCTATTT GGCCAGGCAT 2940  
 TAGCCCAAGA CTTGAGYCAR TYMTCATACC TGGACACTCT TGTCTTCRG TAKGTGGATG 3000  
 ATTTACTTTT RGCYGCCYRT TCAGAAACCT TGTGCCATCA AGCCACCCAA GCRCTCTTMA 3060  
 ATTTCTCGC YACCTGTGGC TACAWGGTTT CCAAACSARA RGCTCARCTC TGCTCACAGC 3120  
 AGGTAAATA CTTAGGRCTA ARATTATCCA AAGGCACCAR GGCCCTCAGT GAGGAAYRYA 3180  
 TCCAGCCTAT ACTGGCTTAT CCTCATCYCA AAACCTTAA GCAACTAAGR GRRTTCCCTG 3240  
 GCRTAAYAGC YTTCTGCCGA AWATGGATTC CCCAGGTWTG GCRAAATAGC CAGGYCATTA 3300  
 WATACASTAA TTAAGGAAAC TCAGAAAGCC AATACCCATT TARTAAGATG GAYAMCTGAA 3360  
 GYMRAAGTGG CTTTCCAGGC CCCTAAAGAA GGCCTTAAAC CCAAGYCCCA GTGTTAAGYT 3420  
 TGCCAACRGG GCAAGACTTT TSTTYATAYR TCACAGAAAA AACAGRAAY AGCTCTRGG 3480  
 GTCCTTACAC AGRTCCRAGG GAYGAGCTTG CAACCYRTGG CRYACCTGAS TAAGGAAAYT 3540  
 GATGTAGTGG CAAAGGGTTG RCYTCAATTG TTAYGGGTAG TGGTGGCAGT AGCAGTYKTA 3600  
 GTATCTGAAG CAGTTAAAAAT AATACAGGGR AGAGATCTTA CTGTGTGGAC ATCTCATGAK 3660  
 GTGAAYRGCA TACTCACTGC TAAAGGAGAC TTGTGGCTGT CAGACAACYG TTTACTTAAA 3720  
 TRTCAGGCTC TATTACTTGA ARGGCCAGTG CTGCRACGTG GCACTTGTGC AACTCTTAAC 3780  
 CCAGYCNCAT TTCTTCCAGA CAATGAAGAA AAGATARAAY ATAAGTGTCA ACAARTAATT 3840  
 TCTCAAACCT ATGCCACTCG AGGGGACCTT YTAGARGTTC CYTTGACTGA TCCYGACCTT 3900  
 CAACTTGTAT ACTGATGGAA GTTCCTTTGT AGAAAAAGGA CTTGAAAAG YGGGGTATGC 3960  
 AGTGGTCAGT GATAATGGAA TAYTTGAAAG TAATCCCTC ACTCCAGGAA CTAGTGCTYA 4020  
 GCTRGAGAA CTAATAGCCY TCAYTKGGGC ACTAGAATTA GGAGAAGRAA AAAGGGYAAA 4080  
 TATATATACA GACTCTRART ATGCTYACCT AGTCNTCCAT GCCCATGMRG CAATATGSAR 4140  
 AGAAAGGGAA TTCCTAACTT CYGAGRGAAC ACCTATCAMA CATCAGGAAG CCATTAGGAR 4200  
 ATTATTAYTG GCWGTACAGA AACCTARAGA GGTGGMAGTC TTACACTGCTY GGGGTCATCA 4260

NAAAGGAAAG RAAAGGGAAA TASAAGRGAA YTGCCAAGCA KATATTGAAG CMAAAAGAGC 4320  
 TGCAAGGCAG GACCCTCCAT TAGAAATGCT TATTAACTT CCCTTAGTAT AGGGTAATCC 4380  
 CTTCCGGGAA ACCAAGCCCC AGTACTCAGC AGGAGAAACA GAATGGGGAA CCTCACGAGG 4440  
 CAGTTTTCTC CCCTCGGGAC GGTTAGCCAC TGAAGAAGGG AAAATACTTT TGCCTGCAAC 4500  
 TATCCAATGG AAATTACTTA AAACCCTTCA TCAAACCTTT CACTTAGGCA TCGATAGCAC 4560  
 CCATCARATG GCCAAATCAT TATTTACTGG ACCAGGCCTT TTCAAACTA TCAAGCARAT 4620  
 AKTCAGGGCC TGTGAAKTGT GCCARARAAA TAATCCCCTG CCTYATCGCC AAGCTCCTTC 4680  
 AGGARAACAA ARAACAGGCC ATTACCCTGR ARAARACTGG CAACTGATTT TACCCACAAG 4740  
 CCCAAACCTC AGGGATTTC A GTATCTACTA GTCTGGGTAR ATACTTTCAC GGGTTGGGCA 4800  
 RAGGCCTTCC CCTGTAGGAC AGAAAAGGCC CAAGAGGTAA TAAAGGCACT AGTTCATGAA 4860  
 ATAATTCCCA GATTCCGACT TCCCCGAGGC TTACAGAGTG ACAATAGCCC TGCTTTCCAG 4920  
 GCCACAGTAA CCCAGGGAGT ATCCCAGGCG TTAGGTATAC GATATCACTT AACTGCGCC 4980  
 TGAAGGCCAC AGTCCTCAGG GAAGGTCGAG AAAATGAATG AAAYACTCAA AGGACATCTA 5040  
 AAAAAGCAAA CCCAGGAAAC CCACCTCACA TGGCCTGYTC TGTTGCCTAT AGCCTTAAAA 5100  
 AGAATCTGCA ACTTTCCCCA AAAAGCAGGA CTTAGCCCAT ACGAAATGCT GTATGGAAGG 5160  
 CCCTTCATAA CCAATGACCT TGTGCTTGAC CCAAGACAGC CAACTTAGTT GCAGACATCA 5220  
 CCTCCTTAGC CAAATATCAA CAAGTTCTTA AAACATTACA AGGAACCTAT CCCTGAGAAG 5280  
 AGGGAAAAGA ACTATTCCAC CCWWTGACA TGGTATTAGT CAAGTCCCTT CYCTCTAATT 5340  
 CCCCATCCCT AGATACATCC TGGGAAGGAC CCTACCCAGT CATTTTATYT ACCCCAAGT 5400  
 CGGTTRAAGT GGCTGGAGTG GAGTCTTGGA TACATCACAC TTGAGTCAA TCCTGGATAC 5460  
 TGCCAAAGGA ACCTGAAAT CCAGGAGACA ACGCTAGCTA TCCTGTGAA CCTCTAGAGG 5520  
 ATTTGCGCCT GCTCTTCAA CAACAACCAG GAGGAAAGTA ACTAAATCA TAAATCCCCC 5580  
 ATGGSCTCC CTTATCATAT TTTTCTCTKT ASTGTTSTTT YACCCTSTTT CACTCTCACT 5640  
 GCACCCCTC CATGCCGCTG TATGACCAGT AGCTCCCCTY ACCMAGAGTT TCTATGGAGA 5700  
 ATGCAGCGTC COGAAATAT TGATGCCCCA TCGTATAGGAG TCTTTSTAAG GGAACCCCC 5760  
 ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATC ACTCTGCCAC TCTTTGCATG 5820  
 CATGCAAATA CTCATTATTG GACAGGAAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 5880  
 GGAGTCACTG TCTGTTGGAC TTACTTCACC CAACTGGTA TGTCTGATGG GGGTGGAGTT 5940  
 CAAGATCAGG CAAGAGAAAA ACATGTAAAA GAAGTAATCT CCCAACTCAC CSGGGTACAT 6000  
 GGCACCTCTA GCCCCTACAA AGGACTAGAT CTCTCAAAC TACATGAAAC CCTCCGTACC 6060  
 CATACTCGCC TGGTAAGCCT ATTTAATACC ACCCTCACTG GGCTCCATGA GGTCTCGGCC 6120  
 CAAAACCCTA CTAAGTGTG GATATGCCTC CCCCTGAAGT TCARGCCATA TGTTTCAATC 6180  
 CCTGTACCTG AACCAATGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 6240  
 GGACCTCTTG TTTCCAATST GGAAATAACC CATACTCAA ACCTCACCTG TGTAAAATTT 6300  
 AGCAATACTA CATAACAAC CAACTCCCAA TGCATCAGGT GGGTAACTCC TCCCACACAA 6360  
 ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCGTTGTTTG 6420

AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCYAT GRCCATCTAC 6480  
 ACTGAACAAG ATTTATACAG TTATGTCATA TCTAAGCCCC GCAACAAAAG AGTACCCATT 6540  
 CTTCTTTTG TTATAGGAGC AGGAGTGCTA GGTGCACTAG GTACTGGCAT TGGCGGTATC 6600  
 ACAACCTCTA CTCAGTTCTA CTACAAACTA TCTCAAGAAC TAAATGGGGA CATGGAACGG 6660  
 GTCGCCGACT CCCTGGTCAC CTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 6720  
 CRAAATCGAA GAGCTTTAGA CTYGCTAACC GCTGARAGAG GGGGAACCTG TTTATTTTTA 6780  
 GGGGAAGAAT GCTGTTATTA TGTTAATCAA TCCGGAATCG TCACTGAGAA AGTTRAAGAA 6840  
 ATTCAGATC GAATACAACG TAKAGCAGAR GAGCTTCGAA ACACTGGACC CTGGGGCCTC 6900  
 CTCAGCCRAT GGATGCCCTG GATTCTCCCC TTCTTAGGAC CTCTAGCAGC TATAATATTG 6960  
 CTACTCCTCT TTGGACCCTG TATCTTTRAC CTCCTTGTTA ACTTTGTCTC TTCCAGAATC 7020  
 GAAGCTGTRA AACTACAAAT GGAGCCCAAG ATGCAAGTCCA AGACTAAGAT CTACCGCAGA 7080  
 CCCCTGGACC GGCCTGYTAG CCCACGATCT GATGTTAATG ACATCAAAGG CACCCCTCCT 7140  
 GAGGAAATCT CAGCTGCACA ACCTCTACTA CGCCCCAATT CAGCAGGAAG CAGTTAGAGC 7200  
 GGTSGTCGGC CAACCTCCCC AACAGCACTT AGGTTTTCCT GTTGAGATGG GGGACTGAGA 7260  
 GACAGGACTA GCTGGATTTC CTAGGCTGAY TAAGAATCCY TAAGCCTAGS TGGGAAGCTG 7320  
 ACCACATCCA CCTTTAAACA CGGGGCTTGC AACTTAGYTC ACACCTGACC AATCAGAGAG 7380  
 CTCCTAAAA TGCTAATTAG GCAAAGACAG GAGGTAAAGA AATAGCCAAT CATYTATTGC 7440  
 MTGAGAGCAC AGCAGGAGGG ACAATGATCG GGATATAAAC CCAAGTYTTC GAGCCGCAAA 7500  
 CGGCAACCCC CTTTGGGTCC CCTCCCTTTG TATGGGAGCT CTGTTTTTCAT GCTATTTTAC 7560  
 TCTATTAAAT CTTGCARCTG CR 7582

(2) INFORMATION FOR SEQ ID NO: 12:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2563 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ACTGCACTCT TCTGGTCCAT GTTGTACG GCTCGAGCTG AGCTTTGCT CGCCATCCAC 60  
CACTGCTGTT TGCCACCGTT GCAGACCCAC TGCTGACTTC CATCCCTCTG GATCTGGCAG 120  
GGTGTCTGCT GTGCTCCTGA TCCAGCGAGG GGCCCATTC CACTCCCAAT CGGGCTAAAG 180  
GCTTGCCATT GTTCCTGCAT GGCTAAGTGC CCAGGTTCAT CCTAATTGAG CTGAACACTA 240  
GTCACTGGGT TCCACAGTTC TCTTCCATGA ACCACGGCTT TTAATAGAGC TATAACACTC 300  
ATCGCAAGGC CCAAGATTCC ATTCCTTGA ATCTGTGAGG CCAAGAACCC TAGGTCAGAG 360  
AACACGAGGC TTGCCACCAT CTGGAAGCA GCCTGCCACC ATCTGGGAAG CGGCCTGCCA 420  
CCATCTTGA AGCCGCCCGC CACCATCTTG GGAGCTCTGG GAGCAAGGAC CTCCCCGCAA 480  
CCCAGTAACA TTAGCGACC ACGAAGGGAC CTCCAAAGCG GTAATATTGG ACCACTTTCA 540  
CTTGCTATTG TGTCCTATCC TTCCTTAGAA TTGGAGGAAA ATACCGGACA CCTGTCGGCC 600  
GGTTAAAAAC GATTAGCGTG GCCTCCGGAC TTAAGAATCA GGTGTGAGGC TATCTGGGGA 660  
AGGGCTTTCT AACAAACCCC AACCRITCTG GGTGGGAAT GTTGGTCTGC CTGGAGCCAG 720  
CTTCCACTTT CAATTTTCCT GGGGAAGCCA AGGGCCGACT AGAGGCAGAA AGCTGTTGTC 780  
CCAAATTCCC GGCAGTAGCC GGTGAGATC ATGGCGCAGC CAGAAGTCTT TACTCCACAG 840  
TCACCCATGC ATGCGCCCT ATCTTTCCTT CTGACCCATA CCTCCTGGGT CCTAACCATG 900  
ACTTTCTTAA AAGGGTAGCC CCAAAATTCT CCTTACCTCT GAATCTACTT CCTCTGATCC 960  
CTGCCTCCTA GGTGCTAATG GTTCAGACTT TCATTTCTCT TAGCAAGTTG TATYTCCAAA 1020  
GGGATATAAG GAAGCTCTAC ACTGTATCCT TAGGCATCTA GGCTCTAAAC CCAGGGAGTC 1080

TTGTCCCTGA TGTCCCAACC GATTTAGGTA TATAGTTCTC GACATGGGCA GTTATGTGGG 1140  
ACCCATTCCC CACCACCCTT GCCAGGGCCC CAAGTTTGTA AATGGCTAAG AGAGGAAAGT 1200  
GAGAGAGAGA GAGACAGAGT GAGACACAGA GAGAGGGAGA GACAGAGAGA GAGACAGAGA 1260  
GGAGAGAGAC ACAGAGAGGG GAGAGACACA GAGAGGAGAA GGGGGCAGAG AGACCAAGAG 1320  
GGAGTCYMAG AGAGAGAGAA AGAAGAAGAA ATAGTAGAAA AAAAAGTGTG CCCTATTCCT 1380  
TTAAAAGCCA GGTAAATTT AAAAACCTA TACTTGATAA TTGAAGGTCT TCTCCATGAC 1440  
CCTGTAACAC TCTAATACTA CCTTGTTCTC AGTGTAACA AGGGTGTTAG CCTGAAAACA 1500  
CTGAGACCGC TGACACCCAT AGCTTTCCTA TAAAAATCC TTAACCCAGT AACCCGAGA 1560  
TGGCCCGCAT GCATTCAATC TGTAGTGGCA ACTGCTTGC TAACAAGAAT AAAGTGGAA 1620  
AGTAACTTTT AGAGGAAACC TCATTGTGAG CACACCTCAC CAGTTCAGAA TTATTCTAAG 1680  
TCAAAAAAGC AAAAAGGTAG CTTACTAACT CAAAATCTT AAAGTATGGG GTTATTTTGT 1740  
TAGAAAAAGG TAATTTAACA CTAATCACTG ATAATTCCTT TAACCCAGAA GATTTCCTAA 1800  
CAGGAGATTT AAATCTTAAT TACCATACAA AGGTCTGACC AGACCTAGGA GGAATCCCT 1860  
TCAGTACAGG ATGATAGATG GTTCCTCCCA GGTGAATGAA AAAAAATCA CAATGGGTAT 1920  
TCAGTAATTG ATAGGGAGAC TCTTGTGGAA GCAGAGTTAG AAAAAGTACC TAATAATTGG 1980  
TCTCCCCAAA CCTGCGAGCT GTTGCACTC AGCCAAGCCT TAAAGTACTT CTAGAATCAA 2040  
AAAGATTATC TCAATCCTGA CTCAAAAGGT TACCTACACC CTCTGTGAAA CGAATTTACT 2100  
TAAGAACTGT TTATGGGACT GCATCTTGAT GGGGCAGCTG GGTGTGTCATG AAATACTCAG 2160



GAATGCAGCC TAGCTCTAGG ACTCACCCCT GAGCACAAAG GCAATGTTGG GCATGCTGGT 2220  
AAAGGACCAC TAGAATCCAG CAGTCCGAAC CCTTCTTTG GGTAAAGAA GGCGGGAAAA 2280  
CAGGCGCAGG ACTGCTACAT TGGTAAGCGT AACTAATCCA ATAAGCAGAG GTCCATGGGT 2340  
GGTGACACAC TCTGGAAAGG AATAAGCATT AGRACCATAG AGGACGCTCT ACGACTAATG 2400  
CTCGTCGGAA AATGACTAGA GGTGCTGGCA TCCCTATGTT CTTTTTTCAG ATGGGAAATG 2460  
TTCCCCCTCA AGGCAAAAAC ACCCCTAAGA TGTATTCTGG ACAATTGGGA CCAATTTGAC 2520  
CCTCAGACTC TAAGAAAGAA ACGACTTATA TTCTTCTGCA GTG 2563

(2) INFORMATION FOR SEQ ID NO: 13:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2585 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TCAGGGATAG CCCCCATCTA TTTGGCCAGG TATTAGCCCA AGACTTGAGC CAGTTCTCAT 60  
ACTTGACAC TCTTGTCTT TGGTATGTGG ATGATCTACT TTTAGCCACC TGTTAGAAA 120  
CCTTGTGCCA TCAAGCCAAC CAAGTGCTCT TAAACTTCCT CGCCACCTGT GGCTACAAGG 180  
TTTCCAAACC AGAGGCTCAG CTCTGCTTAC AGCAGGTTAA ATACTTAGGG CTAAAATTAT 240

CCAAAGGCAC CAGGGCCCTC AGTGAGGAAC GTATCCAGCC TATACTGGCT TATCCTCATC 300  
CCAAAACCCCT GAAGCAATTA AGAGGGTTCC TTGGCATAAA AGGCTGCTGT TGAATATGGA 360  
TTCCCAGGTA CAATGAAATA GCCAGGCCAT TATACACACT AATTACGGGA ACTCAGAAAG 420  
CCAATACCCA TTTAGTAGAA TGGACACCTG AAGCAGAAGC GGCTTTCCAG GCCCTAAAGA 480  
AGGCCCTAAT CCAAGCCCCA GTGTTAAGCT TGCCAATGGA GCAAGACTTT TCTTTATATG 540  
TCACAGAAAA AAAACAGGA ATAGCTCTAG AAGTCCTTAC ACAGGTCCGA GGGACCAGCT 600  
TACAACACAT GGCATACCTG AGTAAGGAAA CTGATGTAGT GGCAAAGGGT TGGACTCATT 660  
GTTTACAGGT AGTGGCAGCA GTAGCAGTCT TAGCATCTGA AGCAGTTAAA ATGATACAGG 720  
GAAGANATCT TACTGTGTGG ACATCTCATG ATGTGAACGG CATACTCACT GCTAAAGGAG 780  
ACTGTGGCTG TCAGACAACC ATTTGCTTAA ATATCAGGCT CTATCACTTG AANGGCCAGT 840  
GCTGCCACTG TGCACTTGTG CAACTCTTAA CCCACCCACA TTTCTTCCAG ACAATGAAGA 900  
AAAGATAGAA CATAACTGTC AACAAGTGAT TGTTCAAACC TACACCGCTC GAAGGGACCT 960  
TCTAGAGGTT CCCTTGACTG ATCCTGAGCT CAACTTCTAT ACTGATGGAA GTTCCTTTTG 1020  
TAGAAAAAGG ACTTCGAAAG GCGGGTATGC AGTGGCCAGT GATAATGGAA TACTTGAAAG 1080  
TAATCCCTTC ACTCCAGAAA CTAGCATTCA GCTGGCAGAA TTAATAGCCT TCACTTGGGC 1140  
ATTAGAACAC AGGAGAAGGA AAAGGAGTAA ATATATATAC AGACTCCAAG TATGCTTACT 1200  
TAGTCCTCCA TGCCCATGCA GCAATATAGA GAGAAAGCGA ATTCCTAACT TCTGAGGGAA 1260  
CACCTATCAA ACATCAGGAA GCCATTAGGA GATTATTACT GGCTGTACAG AAACCTAGAG 1320

GTGGCAGTCT TACATGGCCG AGATCATCAG AAAGGAAAAG AAAGGGAAAT AGAAGGGAAC 1380  
TGCCAAGTGG ATATTGAAGC CAAAAGAGCT GCAAGGCGGG ACCCTCCATT AGAAATGCTT 1440  
ATAGAAGGAC CCCTAGTACA GGGCAATCCC CTTCAGGAAA CCAAGCCCCA ATACTCAGCA 1500  
GAAGAAATGG AATGGGGAAC CTCATGAGGA CATAGTTTCC TCCCCTCAGG ATGGCTAGCC 1560  
ACCAAAGAAG GAAAAATACT TTTGCCTGCA GCTAACCAAT GGAAATTACT TAAAACCCCTT 1620  
CACCAAACCT TTCGCTTAGG CATTGATAGC ACCCATCAGA TGGCTAAATC ATTATTTACT 1680  
AGACCACACC TTTTCAAAC TATCAAGCAG ACAGTTAGGG CCTGTGAAGT GTGCCAAGA 1740  
AATAATCCCC TGCCTTATCG CCAAACCTCT TCAGGAGAAA AAAGAACAGG CCATTACCCA 1800  
GGAGAAGAGT GGCAACTAGA TTTTACCAC ATGCCCAAAT CTCAGGGATT TCAGTATCTA 1860  
CTAGTCTGGG TAGATACTTT CACTGGTTGG GCGGAGGCCT TCCCTTGTAG GACAGAACAG 1920  
GCCCATGAGG TAATAAAGGC ACTAATTCAT GAAATAATTC CCAGATTGAG ATTTCCCCAA 1980  
GGCTTACAGA GTGATAACGG CCCCACTTTC AAGGCTACAG TAACCCAGGG AGTATCCCAG 2040  
ACATTAGACA TACAATATCA CTTACACTGA GCCCGGAGGC CACAATCCTC AGGAAAGTTG 2100  
AGAAAATGAA TGAAACGCTC AAATGACATC TAAAAAGCT AACCTAAGAA ACCCACCTCT 2160  
CATGGTTTGC TCTGTTGCCT ATAGCCTTAG TAAGAATCCG AACTCTCCC CAAAAGCGG 2220  
GACTCAGCCC ATACGAAATG CTGTATGGAC GGCCCTTCCT AACCAATGAC CTTGTGCTTG 2280  
ACCTAGAGAT GGCCAACCTA GTTGCAGATA TCCCTCCTTA GCCAAATATC AACAAGTTCT 2340  
TAAAACGTCA CAGGGAACCT GTCCCTGAGA GGAGGGAAAG GAATTATTCC AACCTGGTGA 2400

CATGGTATTA GTGAAGTCCC TTCCCTCCAA CTCCCCATCC CCTGGATACA TCCTGGGAAG 2460  
GACCCTACTC AGTCATTTTA TCTATCCCAA CCGCGGTAA AATGGCTGGA GTAGAATCTT 2520  
GGATACATCA CATTCGAGTC AAACCCTAGA TACTGCCACA AGGAACCTGA AAATCCAGGA 2580  
GACAA 2585

(2) INFORMATION FOR SEQ ID NO: 14:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2575 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA  
(iii) HYPOTHETICAL: NO
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GGGATAGCCC CCATCTATTT GGCCAGGCAT TAGCCCAAGA CTTGAAGCCA ATTCTCATAC 60  
CTGGACACTC TTCTCCTTTG GTATGTGGAT GATTACTTT TAGCTTCCTG TTCAGAAACC 120  
TTGTGCCATC AAGCCACCCA AGCACTCTTA AATTCCTCG CTACCTGTGG CTACAAGGTT 180  
TCCAAACCAA AGACCCAGCT CTGCTCACAG CAGGTAAAT ACTTGGGGCT AAAATTATCC 240  
AAAGGCACCA GGGCCCTCAG TGAGGAACGT ATCAAGCCTA TACTGGCTTA TCCTCATCCC 300  
CAAATCCTAA AGCAACTAAG AGAGTCCTT AGCATACAG GTTCTGCTG AATATGGATT 360

CCCAGGTATG GCAAAATAGC CAGACCATTA TATACGCTAA TTAAGGAAAC TCAGAAAGCC 420  
AATACCCATT TAGTAAGATG GATACCTGAA GCAGAAGCAG CTTTCCAGGC CCTAAAGAGG 480  
GCCCTAACCC AAGCCCCAGT GTTAAGCTTG CCAACAGGGC AAGACTTTAC TTCGTATGTC 540  
ACAGAAAAAA CAGGAAATAG CTCTAGGAGT CCTTACACAA GTCTGAGGGA TGAGCTTGCA 600  
ACCCATGGCA TACCTGAGTA AGGAAATTGA TGTAAGTGGCA AAGGGTTGGC CTCATTGTTT 660  
ATGGGTAGTG GCGGCAGTAG CAGTCTTAGC ATCTGAAGCA GTTAAAATGA TACAGGGAAG 720  
AGATCTTACT GTGTGGACAT CTCATGATGT GAATGGCATA CTCACTGCTA AAGGAGACTT 780  
GTGGCTGTCA GACAACCATT TACTTAAATA TCAGGCTGTA TTACTTGAAG GGCCAGTGCA 840  
GCAACTGCGC AGTTGTGCAG CTCTTAACCC AGCCACATTT CTTCCAGACA ATGAAGATAG 900  
AACATAACTG CCAACAAGTA ATTTCTCAA CCTAGGCCGC TCGAGGGAAC CTTTTAGAGG 960  
TTCCCTTAAC TGATCCCGAC CTCAACTGT ATACTGATGG AAGTTCCTTT GTAGAAAAG 1020  
GACTTTGAAA AGTGGGGTAT GCAGTGCTCA GTGATAATGG AATACTTGAA AATAATCCCT 1080  
TCATTCCAGG AACCAGCGTT CAGCTGGCAG AATTAATAGC CCTCACTCGG GCATTAGAAT 1140  
TAGGAGAAGG AAAAAGGGTA AATACACATA CAGATTCTAA GTATGTTTAC TTAGTCCTCC 1200  
GTGCCCACGC AGCAATATGG AGAGAAAGGG AATGCTTAAC TTCTGAGGGA ACACCTATCA 1260  
AACATCAGGA AGTTATTAGG AGATTATTAT TGGCTATACA GAAACCTAAA GAGGTGGCAG 1320  
TCTTACACTG CTGGGGTGGT CAGAAAGAAA AGGAAAGGGA AATAAAAGGG AACTGCCAAG 1380  
CGGATATTGA AGCCAAAAGA GCCGCAAGGC AGGACCCTCC ATTAGAAATG CTTATAGAAG 1440

GACCCCTAGT ATGGGGTAAT CCCCTCCGGG AAACCAAGCC CCAATACTTA GAAAAAGAAA 1500  
TAGAATGGGG AACCTCACGA GGACATAGTT TCCTCCCCTC AGGATGGCTA GCCACCGAAG 1560  
AAGGAAAAAT ACTTTTGCCT GCAGCTAACC AATGGAAATT ACTTAAAACC CTTCACCAAA 1620  
CCTTTCACCT AGACATTGAT AGCACCCTATC AGATGGCCAA ATCATTATTT ACTGGACCAG 1680  
GCCTTTTCAA AACTATCAAG CAGCTAGTCA GGGCCTGTGA AGTGTGCCGA AGAAATAATC 1740  
CCATGCCTTA TCACCAAGCT CCTTCAGGAG AACAAAGAAC AGGCCATTAC CCAGGAGAAG 1800  
RVTGGCAACT AGATTTTACC CACATGCCCA AATCTCAGGG ATTTCAGTAT CTACTAGTTT 1860  
GGGTAGATAC TTTCACTGGT TGGGCAGAGA CCTTCCCCTG TAAGACAGAA AAGTCCCAAG 1920  
AGGTAATAAA GGCATTAGTT CATGAAATAA TTCCCAGATT CAGACTTCCC TGAGGCTTAC 1980  
AGAGTGACAA TGGCCCTGCT TTCAAGGCTA CAGTAACCCA GGAGTATCCC AGGTGTTAGG 2040  
TATACAATAT CACTTACACT GCGCCTGGAG GCAGTCCTCA GGAAGGCCG AGAAACTGAA 2100  
TGAAACACTC AAACGACATC TAAAAAAGC TAACCCAGGA AAACCACCTC ACATGGCCTG 2160  
CTCTGTTGCC TATAGCCTTA CTAAGAATCC AAAACTCTCC CCAAAAGCA GGACTTAGCC 2220  
CATACGAAAT GCTATATGGA TAGCCCTTCC TAACCAATGA CCTTGTGCTT GACTGAGAGA 2280  
GAGCCAACTT AGTTGCAGAC ATCACCTCCT TATCCAAATA TCAACAAGTT CTTAAAACAT 2340  
TACAAGGAGC CTGTCCCCGA GAAGAGGGGA AGGAACTATT CCACCCTGGT GACATGGTAT 2400  
TAGTCAAGTC CCTTCCCTCT AATTCTCATT GCCTAGATAT ATCCTGGGAA GGACCCTACC 2460  
CAGTCATTTT ATCTACCCCA ACCGCAGTAA AAGTGCTGG AGTGGAGTCT TGGATACATC 2520  
ACACTCGAGT CAAACCCTGG ATATTACCAA AGGAACCTGA AAATCCAGGA GACAA 2575

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 783 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

66922 "4209460  
TGAGAGACAG GACTAGCTGG ATTCCTAGG CYGACTAAGA ATCCYTAAGC CTAGSTGGGA 60  
AGGTGACCAC RTCCACCTTT AAACACGGGG CTTGCAACTT AGYTCACACC TGACCAATCA 120  
GAGAGCTCAC TAAAATGCTA ATTAGGCAAA GACAGGAGGT AAAGAAATAG CCAATCATYT 180  
ATTGCMTGAG AGCACAGCAG GAGGGACAAY RATCGGGATA TAAACCCARG YHTTCGAGCY 240  
GGCAACRGCA GMCCCCCTTT GGGTCCCYTC CCTTTGTATG GGAGCTCTGT TTTCATGCTA 300  
TTTCACTCTA TTAAATCTTG CARCTGCRCT CTTCTGGTCC ATGTTTCTTA CGGCTYGAGC 360  
TGAGCTTTYG CTCRCRRTCC ACCACTGCTG TTTGCCRCCA CCGCANACCY GCCGCTGACT 420  
CCCATCCCTC TGGATCMTGC AGGGTGTCCG CTGTGCTCCT GATCCAGCGA RGCRCCTT 480  
GCCGCTCCCA ATYGGGCTAA AGGCTTGCCA TTGTNCTGCG AYGGCTAAGT GCCTGGGTTY 540  
RTYCTAATTG AGCTGAACAC TANTCACTGG GTTCCATGGT TCTCTTCTGT GACCCACRGC 600  
TTCTAATAGA RCTATAACAC TYACCRCATG GCCCAAGRTT CCATTCCCTG GAATCCRTRA 660  
RGSCAACGAA CYCCASGTCA GAGAAYACGA RGCTTGCCAC CATCTTGGAA GCGGCCTGCT 720  
ACCATCTTGG AAGTGTTCA CCACCATCTT GGGACCTCTG TGAGCAAGGA CCCCCMRGTR 780  
15 ACA 783

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 20 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

5      **TGTCCGCTGT GCTCCTGATC**      20

(2) INFORMATION FOR SEQ ID NO: 17:

10      (i) SEQUENCE CHARACTERISTICS:  
         (A) LENGTH: 21 base pairs  
         (B) TYPE: nucleotide  
         (C) STRANDEDNESS: single  
         (D) TOPOLOGY: linear

15      (ii) MOLECULE TYPE: DNA

         (iii) HYPOTHETICAL: NO

20      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

**ATGCACTCTG GCTGGGCCAA T**      21

(2) INFORMATION FOR SEQ ID NO: 18:

25      (i) SEQUENCE CHARACTERISTICS:  
         (A) LENGTH: 21 base pairs  
         (B) TYPE: nucleotide  
         (C) STRANDEDNESS: single  
         (D) TOPOLOGY: linear

30      (ii) MOLECULE TYPE: DNA

         (iii) HYPOTHETICAL: NO



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

ACCATTTGAC CCTCAGACAC T

21

5 (2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleotide

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

15 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AACCCTTTGC CACTACATCA ATTT

24

20

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

25 (B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

30

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

35 TCAGGGATAG CCCCCATCTA T

21

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

5 (B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

10

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

15 TTGTCCTCTG GATTTTCAGG TT

22

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 20 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

30

GGACCCTACC CAGTCATTTT

20

(2) INFORMATION FOR SEQ ID NO: 23:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

5

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

10 ATCAGGAGCA CAGCGGACAC

20

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH: 22 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

25

GGACATCCAA AGTGATACAT CC

22

(2) INFORMATION FOR SEQ ID NO: 25:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

5     **AATGTATGGC CTGAAGTGCA G**     21

(2) INFORMATION FOR SEQ ID NO: 26:

10     (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 22 base pairs  
       (B) TYPE: nucleotide  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: linear

15     (ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

20     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

**CTTCCCAGGA TGTATCACTT TG**     22

(2) INFORMATION FOR SEQ ID NO: 27:

25     (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 24 base pairs  
       (B) TYPE: nucleotide  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: linear

30     (ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

35     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

**CACTGCAGAA GAATATAAGT CGTT**     24

(2) INFORMATION FOR SEQ ID NO: 28:

REPLACEMENT SHEET (RULE 26)

669221 42054450

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleotide  
5 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- 10 (iii) HYPOTHETICAL: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GCTTCCAAGA TGGTGGCAAG C

21

15

- (2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 678 base pairs  
20 (B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- 25 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TCAGGGATAG CCCCCATCTA TTTGGCCAGG CATTAGCCCA AGACTTGAGC CAGTTCTCAT 60  
ACCTGGATAT TCTTGTCTT TGGTATGCGG ATGATTTACT TTTAGCCGCC CGTTCAGAAA 120  
CCTTGTGCCA TCAAGCCACC CAAGTGCTCT TAAATTTCTT CGCCACCTGT GGCTACAAGG 180  
TTTCCAAACC AAAGGCTCAG CTCTGCTCAC AGCAGAAGGC TATTTACCCT AAATACTTAG 240  
GGCTGAAATT ATCCAAAGGC ACCAGGGCCC TCACTGAGGA ATGTATCCAG CCTATACTGG 300  
CTTATCCTTA TCCCAAACC CTAAACAAC TAAGAAGGT CTTGGCATA ATAGGCATAA 360  
CAGGCATAAC AGGTTTCTGC TGAATATGGA TTCCCAAGTA CGGCAAATA GCCAGACCAT 420  
TATATACACT AATTAAGGAA ACTCAGAAAG CCAATACCCA TTTAGTAAGA TGGACACCTG 480  
AAGCAGAGGC AGCTTTCCAG GCCGTAAAGA ACACCCTAAC CCAAGCCCCA GTGTTAAGCT 540  
TGCCAGCGGG GCAAGACTTT TCTTTCTGTG TCACAGAAAA AATAGGAATA GCTNTAGGAG 600  
TCCTTACACA GGTCCGAGGG ACCAGCTTGC AACCCATGGC ATACCTGAGT AAGGAAATTG 660  
5 ATGTAGTGGC AAAGGGTT 678

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:  
10 (A) LENGTH: 536 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
15 (ii) MOLECULE TYPE: DNA  
(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCAATCTCCA TGTGTATCC CCTTCCCCAA CTAATAAGGA CCCCCCTTC AACCCAAACA 60  
GTCCAAAAGG ACATAGACAA AGGAGTAAAC AATGAACCA AGAGTGCCAA TATTCCTGG 120  
TTATGCACCC TCCAAGCGGT GGGAGAAGAA TTCGGCCCAG CCAGAGTGCA TGTACCTTTT 180  
TCTCTCTCAC ACTTGAAGCA AATTAAATA GACCTAGGTA AATTCTCAGA TAGCCCTGAT 240  
GGCTATATTG ATGTTTTACA AGGATTAGGA CAATCCTTTG ATCTGACATG GAGAGATATA 300  
ATATTACTGC TAAATCAGAC GCTAACCTCA AATGAGAGAA GTGCTGCCAT AACTGGAGCC 360  
CGAGAGTTTG GCAATCTCTG GTATCTCAGT CAGGTCAATG ATAGGATGAC AACGGAGGAA 420  
AGAGAACGAT TCCCCACAGG GCAGCAGGCA GTTCCCAGTG TAGCTCCTCA TTGGGACACA 480  
GAATCAGAAC ATGGAGATTG GTGCCGCAGA CATTTAAAGC TTTCCCCGGG TACCGA 536

5 (2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 591 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA

15

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCATGGCCAT CTACACTGAA CAAGATTTAT ACAATCATGT CGTACCTAAG CCCCACAACA 60  
AAAGAGTACC CATTCTTCCT TTTGTTATCA GAGCAGGAGT GCTAGGCAGA CTAGGTACTG 120  
GCATTGGCAG TATCACAACC TCTACTCAGT TCTACTACAA ACTATCTCAA GAAATAAATG 180  
GTGACATGGA ACAGGTCACT GACTCCCTGG TCACCTTGCA AGATCAACTT AACTCCCTAG 240  
CAGCAGTAGT CCTTCAAAAT CGAAGAGCTT TAGACTTGCT AACCGCCAAA AGAGGGGGAA 300  
CCTGTTTATT TTTAGGAGAA GAACGCTGTT ATTATGTAA TCAATCCAGA ATTGTCACTG 360  
AGAAAGTTAA AGAAATTCCA GATCGAATAC AATGTAGAGC AGAGGAGCTT CAAAACACCG 420  
AACGCTGGGG CCTCCTCAGC CAATGGATGC CCTGGGTCT CCCCTTCTTA GGACCTCTAG 480  
CAGCTCTAAT ATTGTTACTC CTCTTTGGAC CCTGTATCTT TAACCTCCTT GTTAAGTTTG 540  
TCTCTTCCAG AATTGAAGCT GTAAAGCTAC AGATGGTCTT ACAAATCTAG A 591

5 (2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 364 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

15 (iii) HYPOTHETICAL: NO



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CTAACCTGAG GATCCAGCAG CAGGACTGAG GGTGCCCCGGG GCAAGTGCCA GCCCATGCCA 60  
TCACCCCTCAG AGCCCCGGGT ATGTTTGACC ATTGAGAGCC AGGAAGTTAA CTGTCTCCTG 120  
GACACTGGCG CAGCCTTCTC AGTCTTACTT TCCTGTCCCA GACAATTGTC CTCCAGATCT 180  
GTCACTATCC GAGGGGTCCT AGGACAGCCA GTCACTACAT ACTTCTCTCA GCCACTAAGT 240  
TGTGACTGGG GAACTTTACT CTTTTCACAT GCTTTTCTAA TTATGCCTGA AAGCCCCACT 300  
CCCTTGTTAG GGAGAGACAT TTTAGCAAAA GCAGGGGCCA TTATACACCT GAACAAGCTT 360  
GAAA 364

5 (2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 538 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

15 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Met	Gly	Leu	Pro	Tyr	His	Ile	Phe	Leu	Cys	Ser	Val	Leu	Ser	Pro	Cys
1				5					10					15	
Phe	Thr	Leu	Thr	Ala	Pro	Pro	Pro	Cys	Arg	Cys	Met	Thr	Ser	Ser	Ser
			20					25					30		
Pro	His	Pro	Glu	Phe	Leu	Trp	Arg	Met	Gln	Arg	Pro	Gly	Asn	Ile	Asp
		35					40					45			
Ala	Pro	Ser	Tyr	Arg	Ser	Leu	Ser	Lys	Gly	Thr	Pro	Thr	Phe	Thr	Ala
	50					55					60				
His	Thr	His	Met	Pro	Arg	Asn	Cys	Tyr	His	Ser	Ala	Thr	Leu	Cys	Met
65					70					75				80	
His	Ala	Asn	Thr	His	Tyr	Trp	Thr	Gly	Lys	Met	Ile	Asn	Pro	Ser	Cys
				85					90					95	
Pro	Gly	Gly	Leu	Gly	Val	Thr	Val	Cys	Trp	Thr	Tyr	Phe	Thr	Gln	Thr
			100					105						110	
Gly	Met	Ser	Asp	Gly	Gly	Gly	Val	Gln	Asp	Gln	Ala	Arg	Glu	Lys	His
		115					120					125			
Val	Lys	Glu	Val	Ile	Ser	Gln	Leu	Thr	Gly	Val	His	Gly	Thr	Ser	Ser
		130					135					140			
Pro	Tyr	Lys	Gly	Leu	Asp	Leu	Ser	Lys	Leu	His	Glu	Thr	Leu	Arg	Thr
145					150					155				160	

His Thr Arg Leu Val Ser Leu Phe Asn Thr Thr Leu Thr Gly Leu His			
	165	170	175
Glu Val Ser Ala Gln Asn Pro Thr Asn Cys Trp Ile Cys Leu Pro Leu			
	180	185	190
Asn Phe Arg Pro Tyr Val Ser Ile Pro Val Pro Glu Gln Trp Asn Asn			
	195	200	205
Phe Ser Thr Glu Ile Asn Thr Thr Ser Val Leu Val Gly Pro Leu Val			
	210	215	220
Ser Asn Val Glu Ile Thr His Thr Ser Asn Leu Thr Cys Val Lys Phe			
	225	230	235
Ser Asn Thr Thr Tyr Thr Thr Asn Ser Gln Cys Ile Arg Trp Val Thr			
	245	250	255
Pro Pro Thr Gln Ile Val Cys Leu Pro Ser Gly Ile Phe Phe Val Cys			
	260	265	270
Gly Thr Ser Ala Tyr Arg Cys Leu Asn Gly Ser Ser Glu Ser Met Cys			
	275	280	285
Phe Leu Ser Phe Leu Val Pro Pro Met Thr Ile Tyr Thr Glu Gln Asp			
	290	295	300
Leu Tyr Ser Tyr Val Ile Ser Lys Pro Arg Asn Lys Arg Val Pro Ile			
	305	310	315
Leu Pro Phe Val Ile Gly Ala Gly Val Leu Gly Ala Leu Gly Thr Gly			
	325	330	335
Ile Gly Gly Ile Thr Thr Ser Thr Gln Phe Tyr Tyr Lys Leu Ser Gln			
	340	345	350

Glu Leu Asn Gly Asp Met Glu Arg Val Ala Asp Ser Leu Val Thr Leu  
355 360 365

Gln Asp Gln Leu Asn Ser Leu Ala Ala Val Val Leu Arg Asn Arg Arg  
370 375 380

Ala Leu Asp Leu Leu Thr Ala Glu Arg Gly Gly Thr Cys Leu Phe Leu  
385 390 395 400

Gly Glu Glu Cys Cys Tyr Tyr Val Asn Gln Ser Gly Ile Val Thr Glu  
405 410 415

Lys Val Glu Glu Ile Pro Asp Arg Ile Gln Arg Ile Ala Glu Glu Leu  
420 425 430

Arg Asn Thr Gly Pro Trp Gly Leu Leu Ser Arg Trp Met Pro Trp Ile  
435 440 445

Leu Pro Phe Leu Gly Pro Leu Ala Ala Ile Ile Leu Leu Leu Leu Phe  
450 455 460

Gly Pro Cys Ile Phe Asp Leu Leu Val Asn Phe Val Ser Ser Arg Ile  
465 470 475 480

Glu Ala Val Lys Leu Gln Met Glu Pro Lys Met Gln Ser Lys Thr Lys  
485 490 495

Ile Tyr Arg Arg Pro Leu Asp Arg Pro Ala Ser Pro Arg Ser Asp Val  
500 505 510

Asn Asp Ile Lys Gly Thr Pro Pro Glu Glu Ile Ser Ala Ala Gln Pro  
515 520 525

Leu Leu Arg Pro Asn Ser Ala Gly Ser Ser  
530 535

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 52 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

10

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Met Glu Pro Lys Met Gln Ser Lys Thr Lys Ile Tyr Arg Arg Pro Leu  
1 5 10 15

Asp Arg Pro Ala Ser Pro Arg Ser Asp Val Asn Asp Ile Lys Gly Thr  
20 25 30

Pro Pro Glu Glu Ile Ser Ala Ala Gln Pro Leu Leu Arg Pro Asn Ser  
35 40 45

Ala Gly Ser Ser  
50

15

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 48 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

Met	Leu	Met	Thr	Ser	Lys	Ala	Pro	Leu	Leu	Arg	Lys	Ser	Gln	Leu	His
1				5					10					15	
Asn	Leu	Tyr	Tyr	Ala	Pro	Ile	Gln	Gln	Glu	Ala	Val	Arg	Ala	Val	Val
			20					25						30	
Gly	Gln	Pro	Pro	Gln	Gln	His	Leu	Gly	Phe	Pro	Val	Glu	Met	Gly	Asp
			35					40						45	